Revisiting probabilistic neural networks: a comparative study with support vector machines and the microhabitat suitability for the Eastern Iberian chub (*Squalius valentinus*)

Rafael Muñoz-Mas[1]*, Shinji Fukuda[2], Javier Pórtoles[3], Francisco Martínez-Capel[1]

[1] Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València, C/ Paranimf 1 – Grau de Gandia 46730, País Valencià, (Spain).

[2] Institute of Agriculture, Tokyo University of Agriculture and Technology, Saiwai-cho 3-5-8, Fuchu, Tokyo 183-8509, (Japan).

[3] Fundación para la Investigación del Clima, C/ Tremps 11, Madrid 28040, (Spain).

*Correspondence to: Rafael Muñoz-Mas, e-mail: rafa.m.mas@gmail.com, voice: +34 622098521

Keywords

differential evolution

habitat suitability model

Iberian Peninsula

machine learning

partial dependence plot

species distribution model

Abstract

Probabilistic Neural Networks (PNNs) and Support Vector Machines (SVMs) are flexible classification techniques suited to render trustworthy species distribution and habitat suitability models. Although several alternatives to improve PNNs' reliability and performance and/or to reduce computational costs exist, PNNs are currently not well recognised as SVMs because the SVMs were compared with standard PNNs. To rule out this idea, the microhabitat suitability for the Eastern Iberian chub (*Squalius valentinus* Doadrio & Carmona, 2006) was modelled with SVMs and four types of PNNs (homoscedastic, heteroscedastic, cluster and enhanced PNNs); all of them optimised with differential evolution. The fitness function and several performance criteria (correctly classified instances, true skill statistic, specificity and sensitivity) and partial dependence plots were used to assess respectively the performance and reliability of each habitat suitability model. Heteroscedastic and enhanced PNNs achieved the highest performance in every index but specificity. However, these two PNNs rendered ecologically unreliable partial dependence plots. Conversely, homoscedastic and cluster PNNs rendered ecologically reliable partial dependence plots. Thus, Eastern Iberian chub proved to be a eurytopic species, presenting the highest suitability in microhabitats with cover present, low flow velocity (approx. 0.3 m/s), intermediate depth (approx. 0.6 m) and fine gravel (64–256 mm). PNNs outperformed SVMs; thus, based on the results of the cluster PNN, which also showed high values of the performance criteria, we would advocate a combination of approaches (e.g., cluster &

heteroscedastic or cluster & enhanced PNNs) to balance the trade-off between accuracy and reliability of habitat suitability models.

1 Introduction

Humans have facilitated species extinctions, invasions, increased soil erosion, altered fire frequency and hydrology, and incited profound changes in primary productivity and other key biogeochemical and ecosystems processes (Ellis et al., 2010). Therefore, in the face of this global change, forecasting future ecosystem states, such as future species geographic distributions or land-use patterns, is currently a central priority in biogeographical and ecological sciences (Eberenz et al., 2016; Evans et al., 2016). As a consequence of this priority, scientists, conservationists and managers are repeatedly compelled to confront new problems requiring data analysis (LaDeau et al., 2016).

Data analysis is largely classified into two broad categories: unsupervised and supervised (Olden et al., 2008). The former focus on revealing patterns and structures in data (e.g., finding groups of co-occurring species), such as the renowned Self-Organising Maps (SOM) (Kohonen, 1982) or the laureate t-Distributed Stochastic Neighbour Embedding (t-SNE) (Van Der Maaten and Hinton, 2008). Conversely, supervised approaches, such as decision trees (e.g., CART; Breiman et al., 1984) or the Generalised Additive Models (GAMs) (Hastie and Tibshirani, 1990), attempt to model the relationship between a set of inputs and known outputs (Olden et al., 2008). Based on the nature of the outputs, supervised learning is likewise classified into two main groups. Therefore, if the outputs are continuous, the supervised technique is used to perform regression (e.g., M5; Quinlan, 1992), whereas for categorical outputs the task is termed classification, although intermediate approaches exist (i.e., ordinal regression) (Gutierrez et al., 2016).

Currently, a number of different approaches to perform classification are available: from the simple *k*-nearest algorithm, which assigns an object to the most common class among a *k* number of neighbours; to the complex deep neural networks that have won numerous contests in pattern recognition thanks to their structure, which consists of a vast number of interconnected neurons disposed in multiple layers (Schmidhuber, 2015). Recent applications in ecology and species distribution modelling of both extremes of this range can be found within scientific literature (e.g., Abdollahnejad et al., 2017; Chen et al., 2014). However, the most popular approaches (e.g., GAMs) are those of intermediate complexity located in the middle of this broad spectrum of alternatives (e.g., Muñoz-Mas et al., 2016d). A multitude of different model categories coexist under the umbrella-term classification, therefore, classification techniques such as GAMs are considered to be a purely statistic approach, while others, such as artificial neural networks (e.g., Multi-Layer Perceptrons – MLPs) (Werbos, 1982; McCulloch and Pitts, 1943), are included within the machine learning and computer science discipline (Olden et al., 2008).

Currently, machine learning algorithms, which make vast use of techniques from mathematical programming and statistics (Sousa et al., 2013), are routinely used to address classification tasks in almost every area of knowledge (Oliva and Cuevas, 2017). Among them, one prominent classification activity is the development of species distribution models, or habitat suitability models, to explore species ecology and predict their occurrence under different management and climatic scenarios (Bennetsen et al., 2016; Guisan et al., 2013). However, habitat suitability modelling has several characteristics that are not present in other environmental classification tasks that may determine the performance and reliability of the models (Elith and Graham, 2009). Species rarity (Wisz et al., 2008), which usually conditions the number of independent observations of the target organism and the close-related data prevalence (i.e., the proportion of presences in a data set) (Fukuda and De Baets, 2016; Mouton et al., 2009), may eventually compromise the performance of the models. Moreover, the susceptibility of the modelling technique to regularisation (i.e., adequate control of parameter tuning and easy selection of variables to prevent overfitting) (Reineking and Schröder, 2006) can affect the credibility of the classifier. Nonetheless, each modelling technique has its own unique characteristics; thus, despite recent advances in Artificial Intelligence (AI), an optimal technique that can be indiscriminately applied can never be envisaged (Yano, 2016; Crisci et al., 2012).

Since the industrial revolution, worldwide human impacts on landscapes and river systems have intensified significantly (Habersack et al., 2014). Therefore, one area of research that has grown steadily in the last few decades is that of ecohydraulics (Casas-Mulet et al., 2016). Ecohydraulics is principally addressed to study the relationship between hydraulics (e.g., water depth or flow velocity) and biota to perform environmental flow assessments. In accordance, an enormous number of different techniques have been used to develop the necessary habitat suitability models for riparian and aquatic organisms, from fuzzy logic (Mouton et al., 2009; Rüger et al., 2005) to random forests (Vezza et al., 2015; Fukuda et al., 2014). Nevertheless, the aforementioned GAMs and MLPs (R Muñoz-Mas et al., 2016b;

Jowett and Davey, 2007) are well represented, while papers employing multiple techniques can no longer be considered a rarity (R Muñoz-Mas et al., 2016a; Fukuda et al., 2013).

Although environmental flow assessment should focus on the different components of riparian ecosystems (Poff et al., 2010), it has traditionally focused on fish species (Tharme, 2003) because they occupy relatively high trophic levels and a broad set of habitats must typically be present to complete their life cycle (R Muñoz-Mas et al., 2016c). Consequently, they have been considered adequate indicators of in-stream habitat constraints (Lorenz et al., 2013). Furthermore, although freshwater fish can be considered a well-studied group, new species continue to be described (Tierno de Figueroa et al., 2013). Therefore, 79 new species of freshwater fishes, such as the Eastern Iberian chub (*Squalius valentinus* Doadrio & Carmona, 2006), have been described in the Mediterranean basin since 2000 (Tierno de Figueroa et al., 2013). In this region, with a high number of endemisms, 70% of the freshwater fish species are either threatened with extinction or already extinct, which is the highest proportion anywhere in the world (Maceda-Veiga, 2013). Native fish have suffered from multiple and recurrent introductions, particularly since 1850, which has been highlighted as one of the main negative factors affecting their survival (R Muñoz-Mas et al., 2016d; Tricarico, 2012). In accordance, within the Mediterranean basin, new habitat suitability models are continuously being developed, both for the invasive and the threatened native species e.g., (e.g., Muñoz-Mas et al., 2016e, 2017; Boavida et al., 2014).

A relatively unknown classification technique within ecological literature in general, and ecohydraulics in particular, are Probabilistic Neural Networks (PNNs) (Specht, 1989, 1990).

PNNs are machine learning classifiers that combine the Bayes theorem for decision-making, which assigns an object to the class that presents the highest value in the corresponding true posterior Probability Density Function (PDF) (e.g., $PDF_{class\ i} > PDF_{class\ j}$), with the Parzen-Rosenblatt window method (Parzen, 1962; Rosenblatt, 1956) to estimate the empirical PDF from a finite data sample (Jin et al., 2002). Although PNNs have been traditionally considered to be a kind of artificial neural network (Bishop, 1995), they differ substantially from other artificial neural networks, such as MLPs; thus, optimising PNNs requires the optimisation of very few parameters (typically only one). This parameter can be set manually (Muñoz-Mas et al., 2014). Therefore, the PNN has been considered to be a one-pass learning approach (Specht, 1990).

PNNs have been proven to be proficient in various tasks, such as: risk assessment (Adeli and Panakkat, 2009), bacterial growth prediction (Hajmeer and Basheer, 2002), fault detection (Chang et al., 2009) or cancer diagnosis (Berrar et al., 2003). Conversely, to the best of our knowledge, there are very few examples of their use in ecology, despite their having demonstrated great performance (Muñoz-Mas et al., 2014; Siira et al., 2009; Corne et al., 2004) and stability over various prevalence datasets (Muñoz-Mas et al., 2014). Nonetheless, the latter is an advantage over other approaches that require case weighting (Platts et al., 2008) or resampling (Allouche et al., 2006). In addition, PNNs have displayed great flexibility in encompassing the hydraulic niche (i.e., discriminating the suitable microhabitats) compared to other approaches that have been considered to be excessively rigid (Muñoz-Mas et al., 2014).

Another machine learning approach that showed great flexibility in general (Belousov et al., 2002), and in particular with determining suitable microhabitats (R Muñoz-Mas et al., 2016a), is that of Support Vector Machines (SVMs) (Vapnik, 1995). Habitat suitability models developed with SVMs proved very accurate when compared with other machine learning classification approaches (R Muñoz-Mas et al., 2016d; Fukuda et al., 2013). As with PNNs, SVMs only require the optimisation of very few parameters (Fukuda and De Baets, 2016; Huang and Wang, 2006). Previous comparisons between PNNs and SVMs typically judged SVMs as the preferable option (e.g., Modaresi and Araghinejad, 2014; Muniz et al., 2010; Öğüt et al., 2009). However, since their inception, PNNs have been the subject of scientific research to improve their performance (Ahmadlou and Adeli, 2010) and/or reduce the computational burden (Kusy and Zajdel, 2015; Miguez et al., 2010; Li and Ma, 2008; Berthold and Diamond, 1998). Consequently, the conclusions of these comparisons may have varied if any of the aforementioned methods to improve PNNs had been employed.

In order to scrutinise the real capabilities of PNNs, we compared four different approaches to develop PNNs with standard SVMs and demonstrated that SVMs do not mandatorily outperform PNNs. The paper is structured as follows: section 2 describes the fundamentals of PNNs and the four different approaches followed to develop PNNs, the theory and settings of SVMs, the optimisation approach for PNNs and SVMs, the training dataset and the comparison performed. In section 3, the accuracy of the four different approaches and the SVM and the reliability of the modelled habitat suitability are presented. In section 4, the results are discussed and integrated with current literature. Finally, the conclusions are provided in section 5.

2 Methods

2.1 Probabilistic Neural Networks – PNNs

Following the precepts of the Bayes theorem, PNNs classify a given input pattern (i.e., a string encompassing one record of each of the *p* input variables) to the class that presents the highest value among the posterior PDFs (Zhong et al., 2007). However, these PDFs are typically unknown (Hajmeer and Basheer, 2002); thus, PNNs circumvent this limitation by employing the Parzen-Rosenblatt window method (Parzen, 1962; Rosenblatt, 1956), or kernel density estimation, to calculate empirical PDFs based on the training patterns (i.e., the strings encompassing, each one, one record of each of the *p* input variables) included in the training dataset (Jin et al., 2002). The main idea behind the Parzen-Rosenblatt method is approximating the PDF by a sum of continuous distribution functions or kernels centred at each training pattern (Adeli and Panakkat, 2009), which have smoothing parameters (σ_j) that control the degree of influence (i.e., the window) of each training pattern towards each coordinate (Fig. 1).

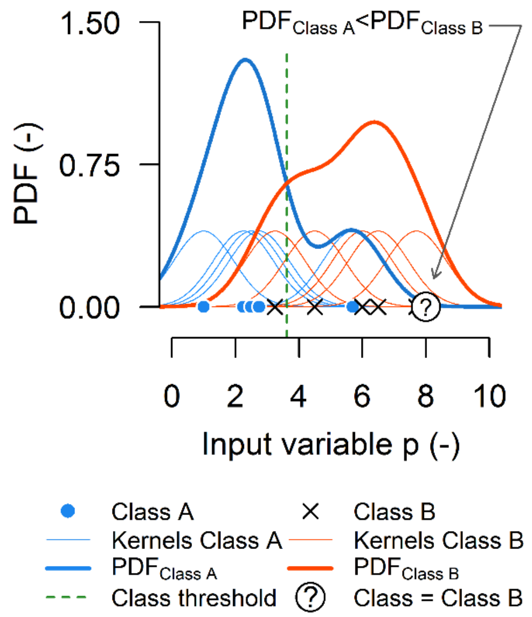

Fig. 1. Example of the classification of an unknown input pattern (?) based on the Bayes theorem and the Parzen-Rosenblatt method to calculate the Probability Density Function (PDF) as the sum of Gaussian kernel functions centred at the training patterns.

Although the kernel function can be chosen from a number of alternatives (e.g., uniform, triangular or Epanechnikov), the bell-shaped normal Gaussian kernel is the most common choice (Kusy and Zajdel, 2015; Modaresi and Araghinejad, 2014; Jin et al., 2002). In accordance, the formula used to calculate the multivariate PDF that combines all the input patterns and variables for each class *m* is:

$$PDF_{class\ m}(x) = \frac{1}{(2\pi)^{p/2} \prod_{j=1}^{p} \sigma_j n} \sum_{i=1}^{n} exp\left[- \sum_{j=1}^{p} \frac{\left(x_j - x_j^i\right)^2}{2\sigma_j^2} \right]; \text{ (Equation 1)}$$

where x is the input pattern to be classified and x_j its j^{th} element (corresponding to the p input variables included in the training dataset), and X^n is the i^{th} training pattern belonging to the class for which the PDF is being calculated, whereas X_j^i corresponds to its j^{th} element (also corresponding to the p input variables) with n equal to the number of training patterns for class *m* (i.e., the total number of training patterns for the class *m*). The σ_j are the window or smoothing parameters, which determine the window of the kernel around the mean of the *p* input variables. Therefore, small values of σ_j produce spiked PDFs with the maxima narrowly centred at the training patterns, whereas large values of σ_j produce smooth PDFs with the maxima instead centred at the region that gathers the maximum number of training patterns (i.e., the region of maximum density of several training patterns) (Fig. 2).

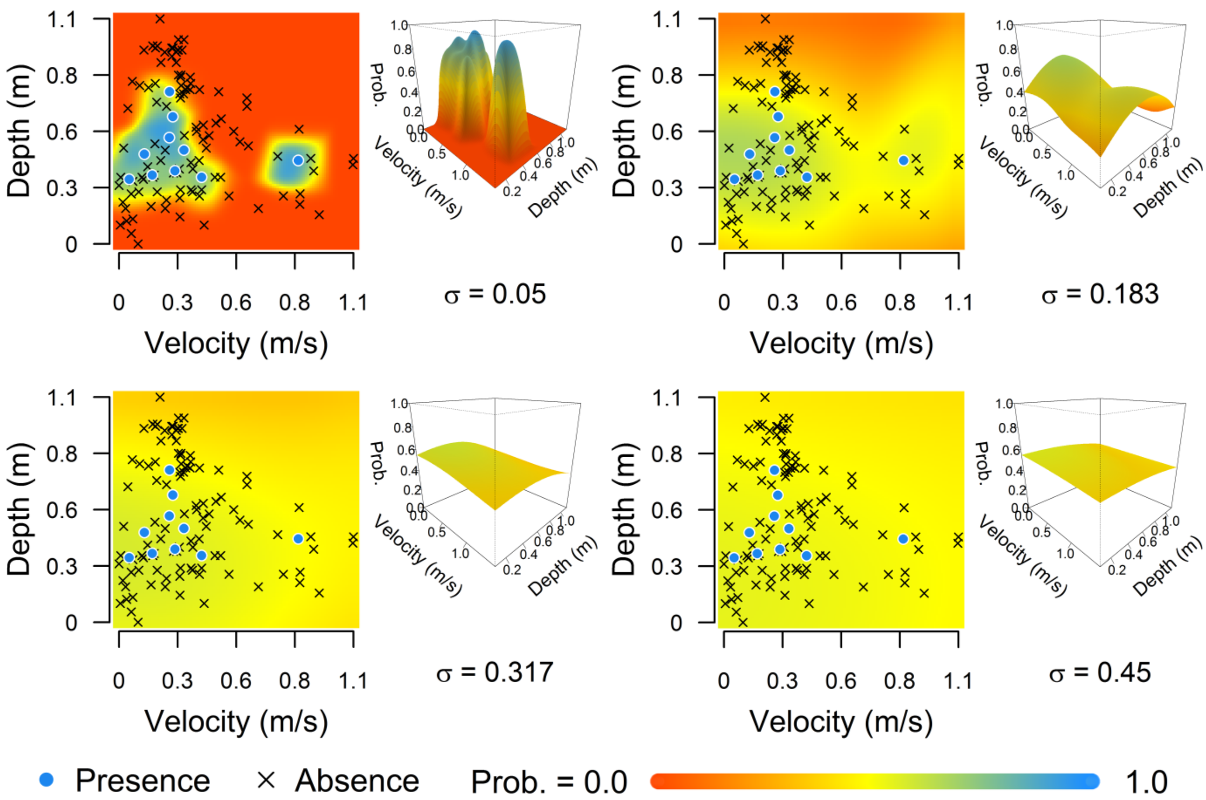

Fig. 2. Effect of the kernel window or smoothing parameter σ on the probability of presence, in a presence-absence classification task, rendered by an homoscedastic Probabilistic Neural Network (PNN) where the smoothing parameter is equalled for each input variables or coordinate (i.e., $\sigma_{Velocity} = \sigma_{Depth}$).

The novelty of the method proposed by Specht (1990, 1989) consisted of breaking up the entire process into a large number of simple processes implemented in a four-layered feed-forward network topology that first calculates the PDFs, following the Parzen-Rosenblatt approach, and then assigns the input pattern to the corresponding class (i.e., solves the inequality of the Bayes theorem) (Fig. 3) (Berrar et al., 2003). The input layer is merely used to supply the input patterns to the pattern layer. The pattern layer has as many nodes as available training patterns and computes the value of each Gaussian kernel function at the input pattern (i.e., at the evaluated point) accounting for the selected smoothing parameters. The summation layer computes the value of the PDF at the input pattern for each class *m*. It is carried out by adding up the outputs of the preceding pattern layer and taking into account the class of the pattern neurons. Consequently, each neuron of this layer is exclusively connected to the pattern neurons corresponding to the same class while the final value is divided by the number of patterns of the corresponding class (*n*) impeding the immediate assignment of the assessed pattern to the outnumbering class (i.e., PNNs are insensitive to data prevalence) (Muñoz-Mas et al., 2014). This is an advantage of PNNs over other machine learning approaches (Muñoz-Mas et al., 2014). Finally, the decision layer compares the values of the PDFs and employs the arguments of the maxima (arg max) to assign the input pattern to the class that presents the highest value among the PDFs. The values of the PDFs are previously standardised by dividing them by the sum of the values of each PDF; thus, the probabilistic outputs are rendered (adding up to one) in addition to the winning class.

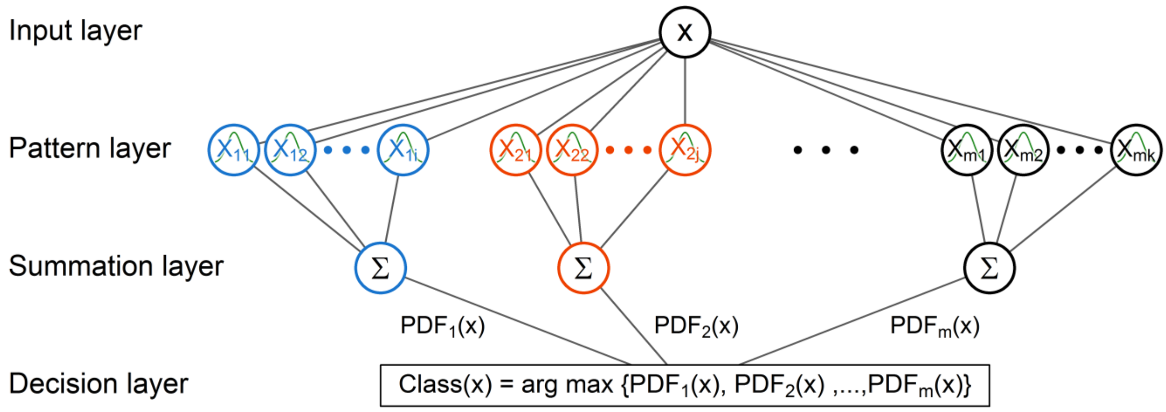

Fig. 3. General architecture of a Probabilistic Neural Network (PNN) where a given input pattern x is classified within a set of m classes. The multi-stemmed architecture is reduced to the two coloured branches depicted on the left when two class problems (e.g., presence-absence) are addressed.

2.1.1 Homoscedastic PNN

This approach is by far the most common way to develop PNN models (e.g., Modaresi and Araghinejad, 2014; Muniz et al., 2010; Öğüt et al., 2009) because it corresponds to the very basic implementation of PNN (Specht, 1989, 1990). The smoothing parameter is equalled for each input variables (i.e., $\sigma_1 = \sigma_2 = \cdots = \sigma_p$); thus, the influence of each pattern on each coordinate coincides (Fig. 4 – Upper left panel). In accordance, a single scalar σ requiring optimisation is used for all pattern neurons (Kusy and Zajdel, 2015). The tested values of the smoothing parameter σ ranged between zero and one (Table 1) (Muñoz-Mas et al., 2014; Muniz et al., 2010).

2.1.2 Heteroscedastic PNN

Although homoscedastic PNNs demonstrated great performance, a single global smoothing parameter σ may be insufficient to achieve the desired accuracy (Chang et al., 2009). By adapting separate smoothing parameters for each coordinate or variable (i.e., $\sigma_1 \neq \sigma_2 \neq$

$\cdots \neq \sigma_p$), the classification accuracy can be greatly improved (Specht and Romsdahl, 1994), as has been corroborated by a number of studies (e.g., Kusy and Zajdel, 2015; Li and Ma, 2008). This type of model is a more elastic classifier, since, in such a case, the influence of each variable on neighbouring points differs (Fig. 4 – Upper right panel) (Kusy and Zajdel, 2015). Thus, the number of smoothing parameters requiring optimisation equalled the number of input variables, and the tested ranges for these parameters (σ_j) also ranged between zero and one (Table 1) (Muñoz-Mas et al., 2014; Verma, 2008).

2.1.3 Cluster PNN

One of the major disadvantage of PNN stems from the fact that it requires one node or neuron for each training pattern, which increases the computational burden (Specht, 1992). Therefore, promptly after their inception, researchers offered various improvements to reduce the number of pattern neurons and hence the computational costs (e.g., Berthold and Diamond, 1998; Burrascano, 1991; Yang and Chen, 1998). Among these improvements, those based on data clustering stood out (Specht, 1992). Nonetheless, depending on the clustering approach, they can be very efficient compared to other approaches that may be inefficient because they need several iterations to converge (e.g., Berthold and Diamond, 1998). Consequently, these approaches rely in a sequential use of unsupervised (clustering) and supervised (PNNs) techniques; thus, cluster PNNs can be homoscedastic, heteroscedastic or enhanced.

Currently, a number of different clustering algorithms have been used as pre-treatments prior to the development of cluster PNNs, such as global k-means (Chang et al., 2009) or j-means (Li and Ma, 2008) (Fig. 4 – Lower left panel). However, over the last 50 years, thousands of clustering algorithms have been published; thus, a number of alternatives are available (Jain, 2010). We advocated the *k*-medoids algorithm (Kaufman and Rousseeuw, 1987) as implemented within the *R* package *cluster* (Maechler et al., 2016). This algorithm clusters data around *k* representative objects or prototypes, named medoids, by minimising the distance between the input patterns and them; thus, it represents a more robust version of *k*-means (Maechler et al., 2016). In addition, they proved to be fast and the actual implementation within the *cluster* package allows one single cluster to be rendered. Therefore, with regard to the example problem of presence-absence, the selection of one single presence pattern and a number of absence patterns surrounding it will fit well the theory around the use of convex hulls (Cornwell et al., 2006) to determine the n-dimensional hypervolume to describe the ecological niche (sensu Hutchinson, 1957). To better illustrate the capabilities of clustering as a pre-treatment, the optimal number of clusters for each class was sought simultaneously with one single smoothing parameter (i.e., homoscedastic PNNs). Therefore, three parameters required optimisation. The maximum number of clusters allowed equalled the maximum number of patterns of the class with the smallest sample size, which for the presence-absence example problem coincided with the sample size of the presence class whereas the single σ ranged between zero and one (Table 1) (Muñoz-Mas et al., 2014; Muniz et al., 2010). The number (#) of clusters was obtained by rounding up the real values ($\mathbb{R}$) given by the optimisation algorithm. Therefore, for the specific example, the *# clusters* were $Class_{presence} = \|\overrightarrow{v_1}\|$ and #

clusters $Class_{absence}$ = $\|\overrightarrow{v_2}\|$ where $\vec{v}$ is the optimal solution, which encodes the best parameters in a vector or chromosome (see below).

2.1.4 Enhanced PNN

In both the aforementioned approaches (homoscedastic and heteroscedastic PNNs) the selected smoothing parameter is used as a global parameter without considering any probable local densities or heterogeneity in the training data (Ahmadlou and Adeli, 2010). To overcome this limitation, a method to improve standard PNNs – named enhanced PNNs – was proposed (Ahmadlou and Adeli, 2010). Enhanced PNNs incorporate local information and existing inhomogeneity, modifying the smoothing parameter of each training pattern in accordance with the proportion of data for the corresponding class within a predefined hypersphere (local circle) of radius *r* (i.e., calculating the proportion of cases of each class and for each pattern below a Euclidean distance *r*) (Fig. 4 – Lower right panel). As a consequence, the smoothing parameter for each training pattern varies as follows:

$$\sigma_{mi} = \alpha_{mi} \times \sigma; \text{ (Equation 2)}$$

where, σ corresponds to the base smoothing parameter, α_{mi} the proportion of training patterns within the local circle for the training pattern *i* that belongs to the class *m*. Finally, σ_{mi} corresponds to the resulting smoothing parameter. In this regard, enhanced PNNs can be viewed as an extension of the heteroscedastic PNNs, where each training pattern presents its own smoothing parameter (Kusy and Zajdel, 2015). However, only two different parameters (σ and r), require optimisation. The smoothing parameter σ ranged between zero and one and the radius of local circles r between zero and two, which for the example problem, with four rescaled variables, coincided with the maximum possible distance between two training patterns (Table 1) (Ahmadlou and Adeli, 2010). The *R* code to implement the four approaches, which is based on the *R* package *pnn* (Chasset, 2013), can be found in Appendix A.

Table 1. Range of the tested parameter settings for the four alternative methods to develop Probabilistic Neural Networks (PNNs) (smoothing parameters σ_j, number of cluster centres and radius of the local circles r) and the Support Vector Machine (SVM) (radial basis kernel function width γ and regularisation parameter C) for the example presence-absence problem.

		Homoscedastic PNN	**Heteroscedastic PNN**	**Cluster PNN**	**Enhanced PNN**	**SVM**
σ_j	**Min.**	0	0	0	0	
	Max.	1	1	1	1	
# clusters	**Min.**			1		
	Max.			Min{$N_{class\ pres.}$, $N_{class\ abs.}$}		
r	**Min.**				0	
	Max.				Max. Euclidean dist.	
γ	**Min.**					0
	Max.					1
C	**Min.**					0
	Max.					500

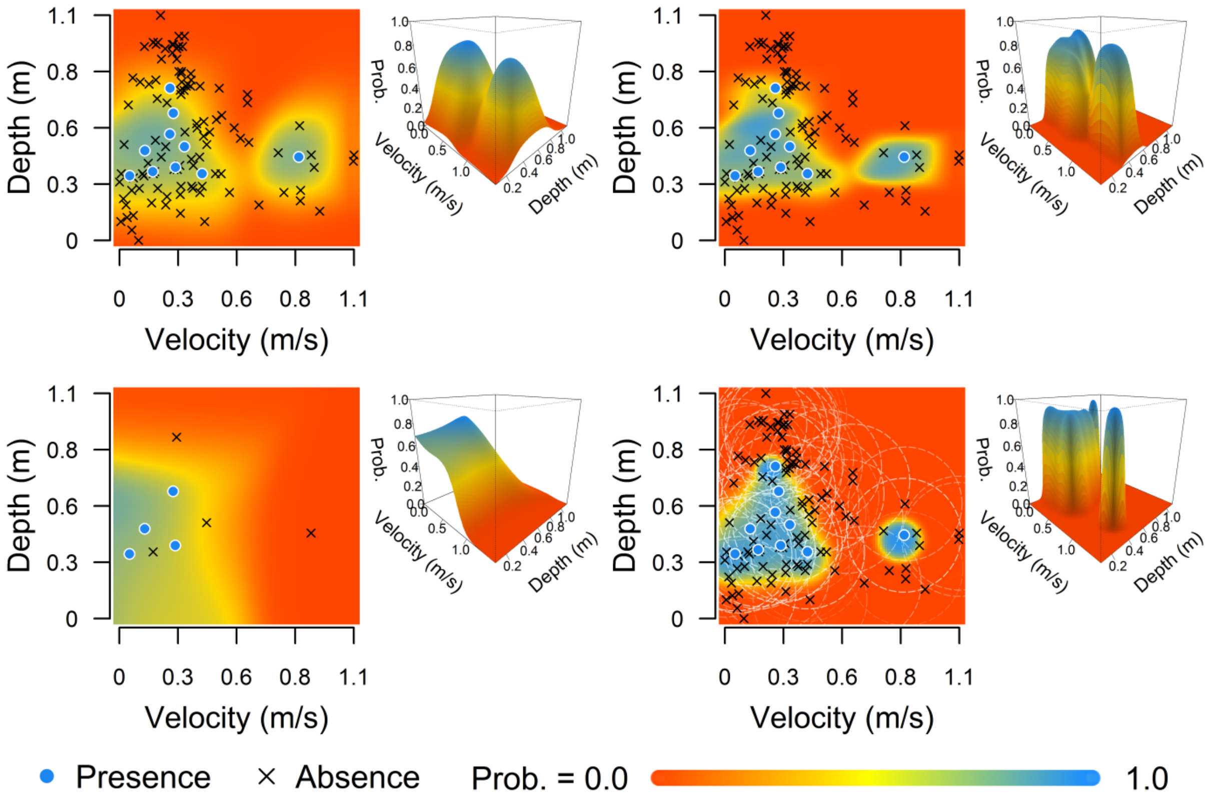

Fig. 4. Example of the differences in the final probability of presence, in a presence-absence classification task, obtained with the four alternative types of Probabilistic Neural Networks (PNNs). The example is based on a random data sample. The upper left panel corresponds to a homoscedastic PNN with one single smoothing parameter (σ), the upper right panel to a heteroscedastic PNN with one smoothing parameter per input variable ($\sigma_{Velocity}$& σ_{Depth}), the lower left panel to a cluster PNN with four clusters per class and one single smoothing parameter (σ) and the lower right panel to an enhanced PNN with local decision circles ($\alpha_i \times \sigma$).

2.2 Support Vector Machines – SVMs

Support Vector Machines (SVMs) is a machine learning approach that employs discriminant hyperplanes to classify the input data (Vapnik, 1995). The fundamental of SVMs, addressed to two-class problems, is to construct an Optimal Separating Hyperplane (OSH), which corresponds to one of the infinite number of existing separating hyperplanes that lie furthermost from both classes, and hence maximises the margin. The margins of the OSH are determined by some cases (i.e., training patterns) which are the so-called support vectors (Fig. 5) (Moguerza and Muñoz, 2006). If the discriminant function between classes is not linear, then the data is projected into a higher-dimensional space (i.e., the feature space) where these data can be linearly separated. This projection is carried out by employing a class of functions called kernels, which perform a nonlinear transformation of the original data. Among these kernel functions, the most popular are polynomial, radial basis and sigmoid (R Muñoz-Mas et al., 2016d; Howley and Madden, 2005). However, in practice, Gaussian Radial Basis Functions (RBFs) have demonstrated to be sufficient to accurately model many real problems (Wu et al., 2012), including habitat suitability modelling (R Muñoz-Mas et al., 2016a; Fukuda et al., 2013). The RBFs exclusively require the optimisation of the kernel width (i.e., the γ parameter), which is related to the variance of the data, and thus determines the radius of influence of samples selected by the model as support vectors. If the problem remains not linearly separable after the kernel transformation, then the misclassified observations can be penalised by a regularisation parameter (C), which defines the trade-off between margin maximisation and error minimisation.

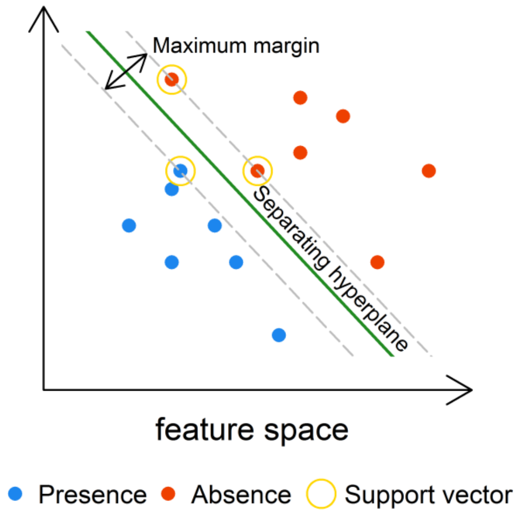

Fig. 5. Optimal Separating Hyperplane (OSH) and selected support vectors in a presence-absence toy example.

The SVMs were developed in *R* (R Core Team, 2015) with the function *svm* implemented within the package *e1071* (Dimitriadou et al., 2011). The selected mapping function was the RBF; thus, the parameters *C* and *γ* required optimisation. The tested ranges of the parameters were based on Huang and Wang (2006) and ranged from 0 to 500 and 0 to 1 for *C* and *γ* respectively (Table 1). Data prevalence affects the performance of SVMs; thus unbalanced datasets may tip the balance towards the outnumbering class (Osuna et al., 1997). In accordance, as with previous studies (R Muñoz-Mas et al., 2016d), each class was weighted by the complementary of its class prevalence (i.e., $1 - prevalence_{class\ m}$). Therefore, the weights used in the example presence-absence problem were 0.06 for the absence class and 0.94 for the presence class (see below for a detailed description of the training dataset).

2.3 Parameter optimisation with Differential Evolution (DE)

Although the parameter requiring optimisation could have been set by employing grid searches (Öğüt et al., 2009), the use of population-based algorithms (e.g., evolutionary and genetic algorithms or particle swarm optimisation) is the most popular approach (Narimani and Narimani, 2013; Miguez et al., 2010; Jin et al., 2002). Therefore, in order to decrease computational cost, a metaheuristic population-based algorithm – the Differential Evolution (DE) algorithm (Storn and Price, 1997) – was used to optimise: the smoothing parameters, the best number of clusters for each class, the radius of the local decision circles and the C and γ parameters of the SVM. The DE algorithm is an evolutionary algorithm inspired by Darwin's process of natural selection and particularly suited to optimise real-valued functions of real-valued parameters (Ardia et al., 2011; Mullen et al., 2011). A user defined number of potential solutions (population) are encoded in vectors (chromosomes or agents) of real-values, and the associated performance (fitness) is calculated for each of these potential solutions (Ardia et al., 2011). Each generation consists of evolving (i.e., creating) a new population from the former population members by mutating and crossing the former population through arithmetic operations, such as addition and subtraction, whose frequency and intensity depends on the parameter settings (Mullen et al., 2011). At each generation, once the entire population has been evolved, only those child vectors that present better fitness substitute their parents (Ardia et al., 2011). The algorithm stops after a specified number of generations, or after the objective function value associated with the best member has been reduced below a specified value (Mullen et al., 2011).

The DE implementation used was that of the *R* package *DEoptim* (Ardia et al., 2011; Mullen et al., 2011) and the optimisation took place following a repeated *k*-fold scheme. Specifically, it followed a three times threefold cross-validation scheme ($3 \times 3_{cross-validation}$) because it proved to be adequate to induce genetically optimised habitat suitability models (Rafael Muñoz-Mas et al., 2016; Stein et al., 2005). In addition, every fold presented similar prevalence to the original dataset (i.e., similar proportion of training patterns for each class). The models were optimised based on a fitness function (equation 3) encompassing several indices arising from the confusion matrix (Table 2) and especially addressed to stimulate over-prediction (Specificity ≤ Sensitivity) (R Muñoz-Mas et al., 2016d, 2016b) because it has been affirmed to be more reliable – from an ecological viewpoint – than under-prediction (Mouton et al., 2010):

$$Fitness = \frac{1}{3\times3}\sum_{i=1}^{3\times3} TSS_i + min\{0, Sn_i - Sp_i\}; \text{(Equation 3)}$$

where *Sn* (*Sensitivity*) corresponds to the ratio of presences correctly classified (i.e., $Sn = \frac{TP}{TP+FN}$), Sp (*Specificity*) corresponds to the ratio of absences correctly classified (i.e., $Sp = \frac{TN}{FP+TN}$) and TSS (True Skill Statistic) to the sum of sensitivity and specificity minus one (i.e., $TSS = Sn + Sp - 1$) (Mouton et al., 2010). In addition, these indices and Correctly Classified Instances or CCI (i.e., $CCI = \frac{TP+TN}{TP+FP+TN+FN}$) were used to evaluate the performance of the different models.

Table 2. Confusion matrix for a two-class problem (e.g., presence-absence). The acronyms correspond to: True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN).

		Observed	
		Presence	**Absence**
Predicted	**Presence**	TP	FP
	Absence	FN	TN

The entire process was fully parallelised employing the 7 cores of an Intel® Core™ i7-4702MQ 2.20GHz with 8GB of RAM while the parameter settings of the optimisation were based on the recommendations described in Mullen et al. (2011) and the package vignette (Table 3), although these parameter settings may also require optimisation to address problems of higher complexity (see e.g., Gibbs et al., 2008). Once the best parameters had been determined, a single model for each alternative approach to develop PNNs and the SVM was trained to perform the subsequent analyses (R Muñoz-Mas et al., 2016d; Fukuda et al., 2013).

Table 3. Differential Evolution (DEOptim) parameter settings. Default values were used in the unlisted arguments.

Operator	Argument name	DEoptim (DE) Setting	Function
Value to be reached	VTR	1	The optimisation stops when this value is achieved.
Evolving strategy	strategy	2	Method employed for mutating and crossing the former population; strategy = 2 corresponds to a uniform mutation operator.
Population size	NP	$10 \times \# parameters$	Number of population members.
Maximum iterations allowed	itermax	$10 \times \# parameters$	Maximum number of generations.
Crossover adaptation	c	0.7	Parameter controlling the crossover. Higher values upweight child vectors.
Relative convergence tolerance	reltol	0.0005	The algorithm stops after *steptol* generations if the absolute improvement of the fitness is lower than *reltol*.
Step tolerance	steptol	$5 \times \# parameters$	See *reltol*.
Crossover probability	CR	0.5	Fraction of the parameter values that are copied from the mutant.

2.4 The training dataset

Although the use of virtual species or in silico datasets is gaining adepts (e.g., Fukuda and De Baets, 2016), we employed a real dataset encompassing the difficulties enumerated in the introduction. The occurrence data for the Eastern Iberian chub were collected at the microhabitat scale (i.e., few m^2 with homogeneous depth, velocity, substrate and cover) during summer low flows (2006) in two separated river stretches of two perennial rivers of the Jucar River Basin District (Fig. 6). The first was located in the Cabriel River (main Jucar River tributary), and the second in the Serpis River.

The Eastern Iberian chub is a small cyprinid (maximum body length = 17.5 cm) (Alcaraz-Hernández et al., 2015) that inhabits the Spanish Levantine region (Perea and Doadrio, 2015). This vulnerable species, whose populations showed marked decreasing trends (IUCN, 2016), occurs principally in streams with clear waters and gravel bottom and prefers moderate flowing stretches (Doadrio and Carmona, 2006).

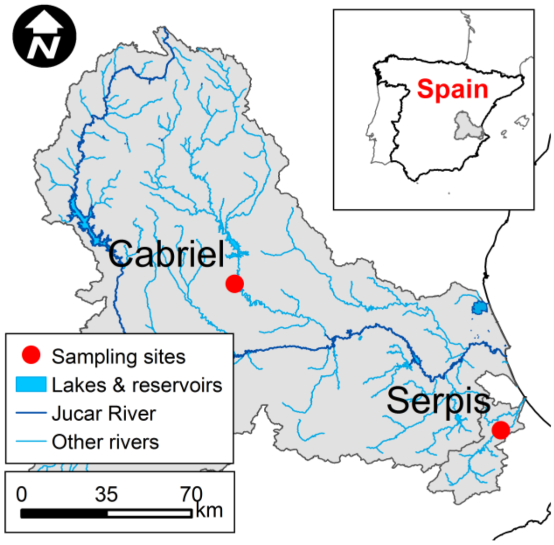

Fig. 6. Location of the sampling sites for Eastern Iberian chub (*Squalius valentinus* Doadrio & Carmona, 2006) within the Cabriel (main Jucar River tributary) and Serpis River basins.

In order to diminish the bias derived from an unbalanced sampling effort over habitat units of fast-flow (e.g. rapid or riffle) and slow-flow (e.g. glide or pool) nature, we selected similar areas (approximately 250 m^2) of these two gross categories (Muñoz-Mas et al., 2012). Following common procedures (see Muñoz-Mas et al., 2014, 2012), the microhabitat study was conducted by underwater observation (snorkelling). The depth, velocity, substrate and cover for the presence data were measured at fish locations, whereas the absence data were collected following a systematic sampling approach (Bovee, 1986). Particularly, these variables were measured in a uniform grid, of approximately 1.5 m^2 per cell, completely covering the sampled habitats. Velocity (m/s) was measured with an electromagnetic current meter (Valeport®, United Kingdom) and depth (m) was measured with a wading rod to the nearest cm. The percentage of each substrate class (i.e., bedrock, boulders, cobbles, gravel, fine gravel, sand, silt and vegetated silt) was visually estimated following the guidelines used in previous studies (e.g., Muñoz-Mas et al., 2012), and these percentages were converted into the substrate index (-) (Mouton et al., 2011) that presents values from zero (vegetated silt) to eight (bedrock). In addition, the presence-absence of cover in the form of: aquatic vegetation, caves, log jams, shade or rocks, was also recorded. These types of cover summarise the concept of structural cover (e.g., boulders, log jams) (Bovee et al., 1998) and escape cover (e.g. vegetation, caves) (Raleigh et al., 1986) and, based on the maximum body size of the species, they were considered present (i.e., one) when any of them occupied an area larger than $0.5 \times 0.5\ m$.

In the end, the Eastern Iberian chub was observed in 40 microhabitats (14 in the Cabriel River and 26 in the Serpis River), whereas the absence data were collected at 607 different microhabitats (304 in the Cabriel River and 303 in the Serpis River). Therefore, the training dataset presented a prevalence of 0.06. The Eastern Iberian chub occurred more frequently at intermediate velocity and depth, when compared with the absence data, it occurred at finer substrates (substrate index = 4) (Fig. 7). Finally, the presence data collected at microhabitats with cover outnumbered the data collected at those with no cover.

The input data involved different units and spanned different ranges, which may lead some variables to dominate the classification (Li and Ma, 2008). Therefore, following previous studies (e.g., Ben-Hur and Weston, 2010; Hajmeer and Basheer, 2002), the input dataset was rescaled between zero and one. This rescaled dataset was used to train the different PNNs and the SVM.

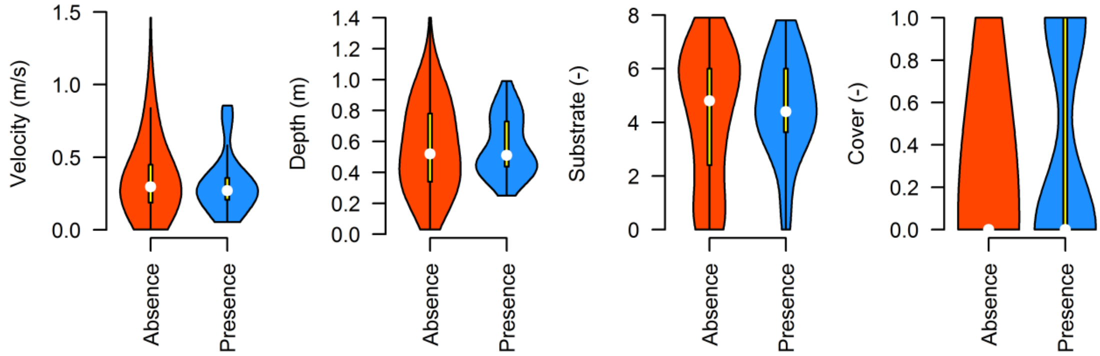

Fig. 7. Violin plots of the data on Eastern Iberian chub (*Squalius valentinus* Doadrio & Carmona, 2006) presence and absence collected in the Cabriel and Serpis Rivers.

2.5 Model evaluation

Developing ecologically reliable habitat suitability models requires balancing the accuracy and complexity of the model via regularisation (Reineking and Schröder, 2006). Otherwise the developed models may violate the ecological gradient theory, which states that species responses to environmental variables are likely to be monotone or unimodal with different degrees of skewness (Austin, 2007).

In order to scrutinise the modelled relationship between the input variables and the probability of presence, partial dependence plots (Friedman, 2001) – based on the code implemented within the package *randomForests* (Liaw and Wiener, 2002) – were developed for each model. Partial dependence plots depict the average of the response variable (i.e., probability of presence) vs a gradient of the inspected predictor variable and accounting for the effects of the remaining variables within the model by averaging their effects (Friedman, 2001). The standard implementation, which consist of first substituting the inspected variable with values along its range to then calculate the mean effect on the response variable (i.e., probability of presence), has been demonstrated to be a useful tool in a number of studies (e.g., Muñoz-Mas et al., 2016b; Shiroyama and Yoshimura, 2016; Vezza et al., 2015). However, averaging the effects of the remaining variables may mask variables interactions (Zurell et al., 2012; Evans et al., 2011). Therefore, instead of one single line chart depicting the mean value vs the gradient of the inspected variable (Ω_z), every 100-quantil or percentile ($Q_z(s)$) was depicted for every value in Ω_z to get a better insight of the modelled habitat suitability. The partial dependence was computed for each of 50 equally spaced values over the range of each examined variable ($m = 50$) except for cover, which is a dichotomous variable ($m = 2$). Therefore, for each predictor (z) within the original training dataset (X) several values along the inspected gradient were calculated following Equation 4:

$$\Omega_z = \left\{ \min(X_z) + \left(\frac{\max(X_z) - \min(X_z)}{m-1} \times k \right) \middle| \, k = 0, \dots, m-1 \right\}; \text{ (Equation 4)}$$

where $\min(X_z)$ and $\min(X_z)$ correspond respectively to the minimum and maximum entries of the inspected variable (X_z) included in X with $z \in \{1, \dots, p\}$ and p equalling the number of predictor variables included in the training dataset (in the example problem $p = 4$). Then, for each value included in Ω_z a modified dataset ($X_z(s)$) is obtained, with $z \in \{1, \dots, p\}$ and $s \in \Omega_z$, by substituting X_z in X by the corresponding value of Ω_z. Then, the resulting dataset $X_z(s)$ is evaluated with the model for which the partial dependence plots are being calculated (i.e., $g(\cdot)$ that can be one of the four different PNNs or the SVM) and the percentiles of the cumulative distribution function $Q_z(s)$, as defined by Gumbel (1939), are calculated ($F(\cdot)$) following Equation 5:

$$Q_z(s) = F(g(X_z(s))); \text{(Equation 5)}$$

3 Results

3.1 Performance

The *DEOptim* algorithm rendered values for the smoothing parameters σ_j for three out of four of the models in a relatively narrow range (i.e., σ from 0.091 to 0.364). Conversely, the parameter for variable velocity in the heteroscedastic PNN presented the smallest ($\sigma_{velocity}$ = 0.030) (Table 4). The latter value contributed to the ecologically unreliable partial dependence plot for this variable (see below).

Table 4. Best parameters obtained for the four different approaches to develop Probabilistic Neural Networks (PNNs) and the Support Vector Machine (SVM).

Model	Optimal parameters
Homoscedastic PNN	σ = 0.106
Heteroscedastic PNN	$\sigma_{velocity}$ = 0.030; σ_{depth} = 0.091; $\sigma_{substrate}$ = 0.211; σ_{cover} = 0.187
Cluster PNN	# $clusters_{presence}$ =22; # $clusters_{absence}$ =20; σ = 0.136
Enhanced PNN	σ = 0.364; r = 0.196
SVM	C = 24.789; γ = 0.163

Enhanced PNN and heteroscedastic PNN achieved the highest values of the fitness function, although the enhanced PNN presented lower variability (Table 5). Regarding the performance criteria used for model evaluation, Heteroscedastic and cluster PNNs presented the best accuracy (CCI) but heteroscedastic PNN presented the highest True Skill Statistic (TSS). Finally, enhanced PNN rendered the best Sensitivity (Sn) whereas the best Specificity (Sp) was obtained with cluster PNN. Looking at the probability of presence rendered in addition to the winning class, the four methods to develop PNN rendered outputs covering the entire feasible output range (i.e., from zero to one) whereas the maximum value obtained with SVM, which alters the classification threshold, was only 0.1 (see also Fig. 9).

The optimisation of the SVM took the shortest time and the heteroscedastic PNN the longest, which was in line with the number of pattern neurons of the PNN and the searching effort, which rose in accordance with the number of optimised parameters (Table 5). The SVM presented the smallest ratio between the number of parameters optimised vs time, whereas among the four approaches to develop PNN, the homoscedastic PNN (tight followed by cluster PNN) presented the smallest ratio and the heteroscedastic PNN the highest.

Table 5. Model performance and confidence interval to evaluate the four different approaches to develop Probabilistic Neural Networks (PNNs) and the Support Vector Machine (SVM): Fitness (Eq. 3), Correctly Classified Instances (CCI), True Skill Statistics (TSS), Sensitivity (Sn), Specificity (Sp) and minimum (Min.) and maximum (Max.) values of the probability of presence obtained during the 3 × 3 cross-validation (nine models) and the lapse of the optimisation (Optimisation was parallelised in an Intel core i7). The best results are in bold.

		Cross-validation						
	Fitness	$\overline{CCI}$	$\overline{TSS}$	$\overline{Sn}$	$\overline{Sp}$	**Min.**	**Max.**	**Time (min)**
Homoscedastic PNN	0.34±0.17	0.65±0.05	0.39±0.10	0.75±0.13	0.64±0.06	**0.0**	**1.0**	6.10
Heteroscedastic PNN	0.45±0.12	**0.68±0.04**	**0.47±0.07**	0.80±0.09	0.67±0.05	**0.0**	**1.0**	73.04
Cluster PNN	0.39±0.14	**0.68±0.04**	0.44±0.08	0.76±0.11	**0.68±0.05**	**0.0**	**1.0**	22.52
Enhanced PNN	**0.45±0.09**	0.65±0.04	0.46±0.08	**0.82±0.09**	0.64±0.05	**0.0**	**1.0**	32.60
SVM	0.33±0.15	0.67±0.04	0.38±0.07	0.71±0.1	0.67±0.05	**0.0**	0.1	**3.85**

3.2 Partial dependence plots

The four tested approaches (Fig. 8) and the SVM (Fig. 9) modelled similar habitat suitability (i.e., similar habitats that would be classified as presence) with the exception of heteroscedastic PNN, which rendered a multimodal mean partial dependence plot for the variable velocity (in black). However, in accordance with the maximum probabilistic values obtained for the SVM (Table 5), the partial dependence plots for this technique presented lower values for all plots (Fig. 9). Nevertheless, SVMs modify the classification threshold; thus, the microhabitats being considered suitable did not change substantially, which maintains the interpretation of these plots.

Setting apart the mean partial dependence plot (in black) for velocity obtained with the heteroscedastic PNN, Eastern Iberian chub presented the highest suitability in microhabitats with low flow velocity (approx. 0.3 m/s) and intermediate depth (approx. 0.6 m), and with mild coarseness – the optimal substrate index was approx. 4, which corresponds to fine gravel (64–256 mm) – and with cover present (i.e., cover = 1). Suitability decreased at the extremes of the range of variables, with the exception of the cover variable. This decrease was especially relevant for high velocity and low depth, and for fine substrate and no cover. Nevertheless, the depicted quantiles revealed that the Eastern Iberian chub is a versatile species that can select a microhabitat when some variables compensate for the low quality of others. In accordance, mean unsuitable conditions at the extremes of the range of variables (in black) presented high values of the probability of presence under particular circumstances (coloured quantiles), which were caused by these infrequent occurrences of the species. As a consequence, some microhabitats with high depth and coarse substrate were evaluated positively, as well as some without cover, in the partial dependence plots of the enhanced PNN.

According to the low $\sigma_{velocity}$, the heteroscedastic PNN exacerbated the mathematical effect of the these infrequent occurrences; thus, the presence of four presence data within 0.730 and 0.856 m/s raised the probability of presence to the maximum. The opposite occurred between 0.467 and 0.730 m/s, where the absence of presence patterns reduced the probability of presence as far as zero. Enhanced PNN also presented relevant irregularities, in spite of the reasonably smooth mean partial dependence plots (in black). It rendered extreme values (i.e., zero and one) in almost every value of the evaluated range. From the mathematical viewpoint, the occurrence of isolated presence data reduced the value of α_{mi} for these patterns and consequently, the PNN rendered extreme values of the PDF when approximating them. On the contrary, cluster PNN reduced such irregularities due to the shrinkage in the number of training patterns, presenting smooth partial dependence plots.

Nevertheless their plots largely coincided with those for the homoscedastic PNN and SVM, although SVM presented the smoothest plots of these three models (Fig. 9).

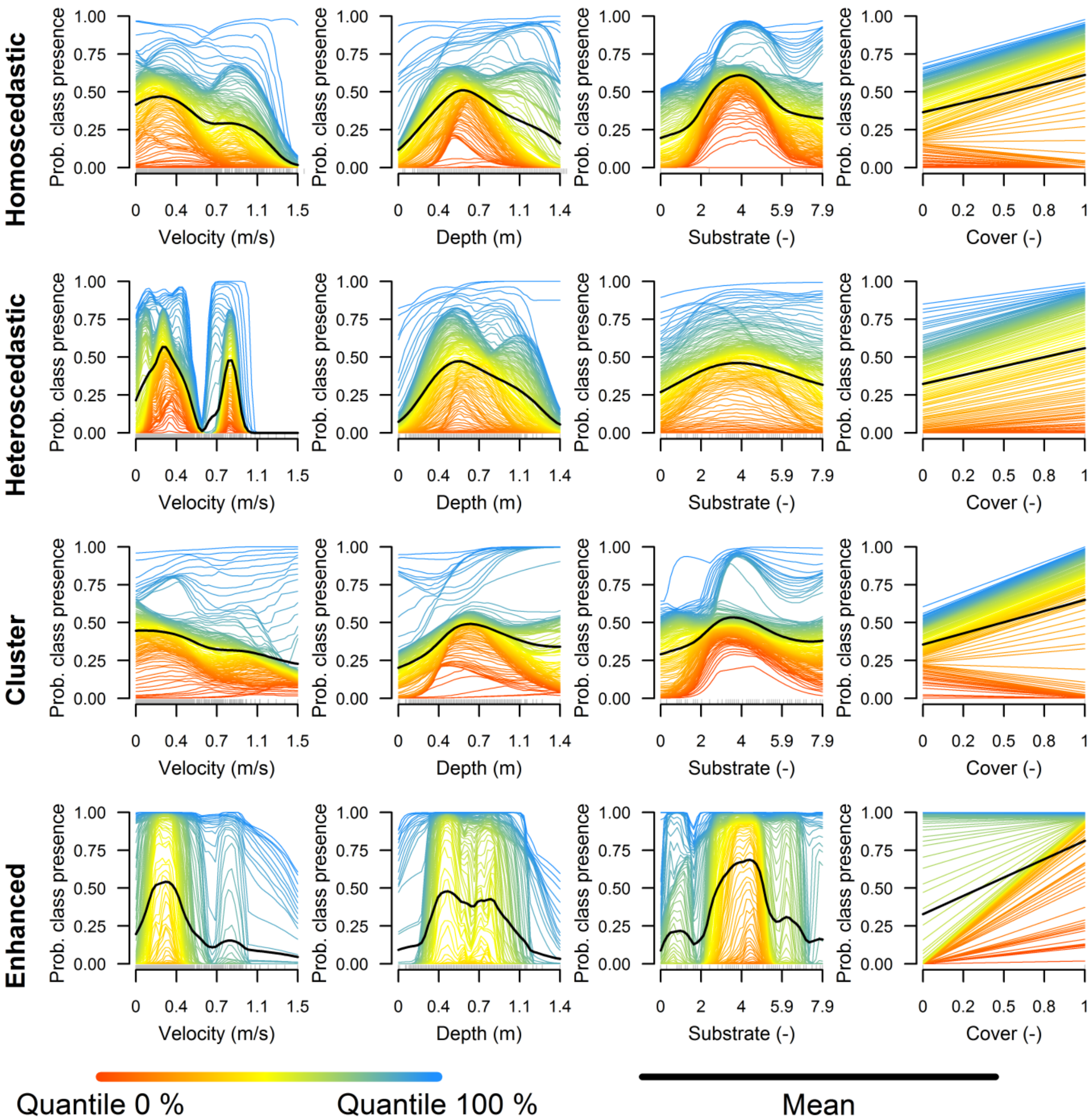

Fig. 8. Partial dependence plots, for the four approaches to develop probabilistic neural networks (homoscedastic, heteroscedastic, cluster and enhanced), depicting the marginal relationship between the suitability (i.e., probability of class presence) and the four microhabitat variables.

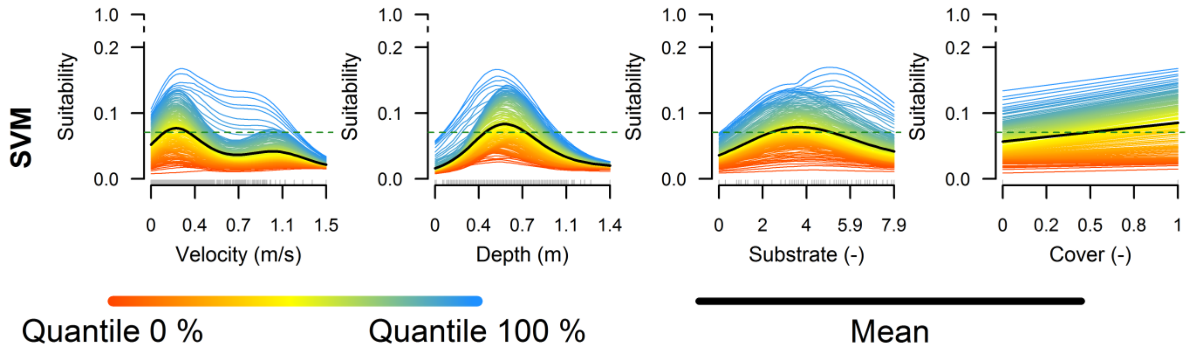

Fig. 9. Partial dependence plots, for the Support Vector Machine (SVM), depicting the marginal relationship between the suitability (i.e., probability of class presence) and the four microhabitat variables. The dashed line depicts the modified classification threshold.

4 Discussion

Four different approaches to develop PNNs have been successfully compared with SVMs demonstrating that SVMs do not outperform every kind of PNN. Nonetheless, the SVM did not present the best value in any of the calculated criteria. In addition, we obtained relevant information about the microhabitat suitability for the Eastern Iberian chub.

4.1 Model characteristics

These four alternative PNNs span the two major contemporary approaches to improve PNNs, which include the smoothing parameter optimisation and pattern neuron reduction (Kusy and Zajdel, 2015; Ahmadlou and Adeli, 2010; Miguez et al., 2010; Li and Ma, 2008).

4.1.1 Smoothing parameter optimisation

Among the group of smoothing parameter optimisations, two of the most common approaches have been tested, although other alternatives exist, such as different smoothing parameters for each class (homoscedastic and heteroscedastic) (Zhong et al., 2007), which can be extended to the extreme of one smoothing parameter per training pattern at the expense of increasing the computational burden (Kusy and Zajdel, 2015). The former could certainly be interesting, although it can be also tackled by employing prior probabilities or uneven misclassification costs (discarded for this study) that may favour the desired class by increasing the final value calculated for the corresponding PDF. In accordance, we considered that the use of different smoothing parameters for each class may require a dedicated study. Regarding the latter, from the ecological viewpoint heteroscedastic PNN may potentially lead to unreliable models (Austin, 2007), although this approach rendered one of the best performances. Therefore, we considered the optimisation of one smoothing parameter per training pattern inappropriate because of the number of parameters and thus increasing risk of overfitting. As a consequence, we would neither advocate heteroscedasticity nor one smoothing parameter per training pattern without the adequate scrutiny of the modelled habitat suitability.

This concern is also extendable for enhanced PNN, although it is indubitable that varying the smoothing parameter to account for local densities and data inhomogeneity improves model performance. Nonetheless, enhanced PNN achieved one of the highest mean values of the fitness function by exclusively optimising two parameters. Enhanced PNNs modify the value of the smoothing parameter based on the proportion of cases within the local circle that are from the same class. However, the use of local circles to account for the data inhomogeneity do not solve the impact caused by the presence of rare or infrequent data, such as the presence data observed between 0.730 and 0.856 m/s, which appears isolated in the input space (see violin plots in Fig. 7). Thus, even the result of a careful cross-validation (e.g., repeated k-fold or leave-one-out) could be biased if the source data set is insufficiently representative as a whole or if a relevant proportion of rare samples are present (Grim and Hora, 2010), resulting in spiked PDFs, as observed in the corresponding partial dependence plots.

With regard to the SVM, the γ parameter of the Gaussian kernel can be seen as the smoothing parameters of the PNNs and, taking into account that they are constant across the input space, the approach followed may resemble that described for the homoscedastic version of the PNN. However, several variants of SVMs exist, which may render different results. Although other authors disregarded this option (Wu et al., 2012), the most basic one consists of varying the kernel function (R Muñoz-Mas et al., 2016d). Nevertheless, with regard to the PNNs, different kernel density functions can also be selected to calculate the PDFs (e.g., triangular, Epanechnikov or Gaussian). Consequently, the number of combinations would be large; thus, we consider that such appealing comparison may also require a dedicated study.

Interestingly, clustering has also been successfully applied as a pre-treatment for SVMs in several ways. For instance, there are studies that tested the impact of different clustering algorithms on accuracy and computational burden (e.g., Wang et al., 2007). Conversely, although methodologically appealing, to the best of our knowledge SVMs did not present any modelling approach analogous to the heteroscedastic and/or the enhanced variants of the PNNs where each axis or each pattern presents its own smoothing parameter σ. Thus, the equivalent to the enhanced approach for an SVM would be the local optimisation of a separate SVM for the neighbourhood of each training pattern. Although this strategy still needs to be the subject of dedicated research to be formally mathematised (e.g., the radii of the neighbourhood may also vary across the input space), it can be ventured that accounting for local densities or inhomogeneity in the training data should also lead to improved outcomes for SVMs. Consequently, we considered the ideas involving the enhanced variant and its stand-alone use, or in combination (e.g., cluster & enhanced PNNs), to be appealing.

On the other hand, the SVM rendered the most ecologically reliable partial dependence plots, although they could be considered deficient if outputs covering the entire feasible range are desired (R Muñoz-Mas et al., 2016a). For instance, SVM will render misleading results of the renowned Weighted Usable Area (WUA) (Bovee *et al.*, 1998). The WUA is calculated for a given flow as the product of the habitat area (e.g., pixels) and habitat suitability of the hydraulics of this area and summed across a river reach (R Muñoz-Mas et al., 2016a). However, given the mathematization of the WUA, a huge amount of low quality habitat may render similar total value as a small area of highly suitable habitat (Person et al., 2014). This is not the case of the Suitable Area (SA), which is the sum of the areas where the models predicted presence (Person et al., 2014). Therefore, the SVM proved only competent to calculate the SA (R Muñoz-Mas et al., 2016a). Previous experiences indicated data overlapping and prevalence as the main causes of this deficiency (R Muñoz-Mas et al., 2016d, 2016a); thus, this study corroborated that Platt's approach for probability calculation is unable to render proper probabilistic results for class-overlapped and low-prevalence datasets (Platt, 2000). Nevertheless, this deficiency could be addressed by employing clustering approaches to balance the data prevalence or alternative routines to train SVMs that are particularly indicated to render reliable probabilistic outputs. Currently, the most promising approach for the latter case is the routine implemented in the *R* package *probsvm* (Zhang et al., 2013), which combines ideas from a number of sources (Shin et al., 2014; Wang et al., 2008; Wu et al., 2004). However, it does not allow case weighting. Consequently, in the end, it may require the use of resampling strategies or balancing algorithms (e.g., SMOTE Chawla et al., 2002) that are unnecessary for PNNs.

4.1.2 Pattern neuron reduction

Clustering as a pre-treatment has been demonstrated to be proficient for the rapid training of accurate PNNs. Nonetheless, taking into account the larger number of fitness function evaluations performed, which was governed by the parameters *NP* and *steptol/reltol* (i.e., $10 \times \# parameters$ and $5 \times \# parameters$), the lapse of the optimisation for cluster PNN should not be considered different than that for the SVM. However, other alternatives exist (e.g., Kusy and Zajdel, 2015; Berthold and Diamond, 1998). Among the group of approaches for pattern neuron reduction one popular approach with a number of examples (e.g., Narimani and Narimani, 2013) is the Dynamic Decay Adjustment (DDA) (Berthold and Diamond, 1998). DDA adds sequentially each training pattern to the PNN, but exclusively retains those patterns that are not redundant and/or in conflict with the remaining classes

(Berthold and Diamond, 1998). Based on the latter, DDA could be more accurate than our cluster approach because it takes into account the distribution of every input class during PNN growth. Nevertheless, it requires several epochs to converge, typically five (Berthold and Diamond, 1998), which may lead to increasing computational costs. On the contrary, compared to the remaining approaches, optimising a different number of cluster centres for each class rendered high performance criteria; cluster PNN rendered the highest accuracy (CCI) and Specificity (Sp). In addition, the partial dependence plots were ecologically reliable because the clustering approach as a pre-treatment reduced the influence of rare data, which typically compromises the reliability of the PNNs (Grim and Hora, 2010; Yang and Chen, 1998). Furthermore, the approach followed to develop the cluster PNN could be combined with heteroscedastic and/or enhanced PNNs (Chang et al., 2008; Yang and Chen, 1998). Therefore, we consider the combination of cluster PNN with other methods to be the most promising approach for ecological studies. Accordingly, we expect these ideas to be the subject of further research.

4.2 Habitat suitability for the Eastern Iberian chub (*Squalius valentinus*)

In spite of the ecologically unreliable partial dependence plots rendered by heteroscedastic and enhanced PNNs, the five models largely converged on the optimal microhabitat for the Eastern Iberian chub. Therefore, the species will preferentially occur in microhabitats with low flow velocity – but not stagnated – of intermediate depth and substrate, and with cover fundamentally present. This description broadly matches the general preferences of the species suggested by Doadrio and Carmona (2006) who stated that the species prefers moderate flowing reaches with clear water and gravel bottom, which in addition may fit the habitat requirements of a number of Iberian species of the genus (e.g., Martelo et al., 2014; Martínez-Capel et al., 2009; Santos and Ferreira, 2008). However, flexible microhabitat use strategies are common among fishes inhabiting Mediterranean streams due to the long-term adaptations to irregular flow regimes (Martelo et al., 2014). Thus, any comparison with others species or studies should be made cautiously, as similarities may be due to particularities, either spatial or temporal.

The optimal values for Eastern Iberian chub showed remarkable similitudes with those obtained for *S. pyrenaicus*, perhaps the closest relative (Doadrio and Carmona, 2006), which was studied in several river reaches of the central Iberian Peninsula (Martínez-Capel et al., 2009). Thus, the only difference was the optimal value of depth for large individuals (> 10 cm), which tended to occupy deeper microhabitats (0.49 to 1.40 m) (Martínez-Capel et al., 2009). Eastern Iberian chub also occurred in microhabitats similar to those occupied by *S. torgalensis*, *S. carolitertii* and *S. aradensis,* all of which have been sampled in other Mediterranean and temperate small streams of the Iberian Peninsula (Martelo et al., 2014; Santos and Ferreira, 2008; Santos et al., 2004). This coincidence was especially relevant for depth, substrate and, to a lesser extent, for cover – which was used less frequently by these species – but the most remarkable difference occurred for velocity (Martelo et al., 2014; Santos and Ferreira, 2008; Santos et al., 2004). However, this is most probably caused by differences in the available microhabitats, which, in these studies, were dominated by shallow and moderate-to-fast flowing riffles and runs (Martelo et al., 2014; Santos and Ferreira, 2008). Taking into account that Eastern Iberian chub rarely exceed 20 cm (Alcaraz-

Hernández et al., 2015; Doadrio and Carmona, 2006) such a discrepancy can be caused by the relatively small size of the species, which may lead to inferior natatorial capacity, as has been demonstrated for other Iberian species (i.e., *S. carolitertii*) (Romão et al., 2012). Nevertheless, while velocity may certainly be a limitation in the occurrence of Eastern Iberian chub, we consider that depth is not. Nonetheless, in Vezza et al. (2015) different results were obtained for what was originally classified as *S. pyrenaicus* based on the morphologic characteristics of the specimens captured in the upper Cabriel that did not match those described in Doadrio and Carmona (2006). However, in light of the information contained in contemporary studies that performed genetic analyses (Perea, 2016 personal communication; Perea and Doadrio, 2015), it is currently suspected that the *Squalius* inhabiting the upper Cabriel is also *S. valentinus*. Therefore, following this supposition, in Vezza et al. (2015) Eastern Iberian chub occurred principally in low gradient and depth (from 1.25 up to 3.5 m) mesohabitats (i.e., pools) of intermediate granularity whereas the presence of macrophytes (one of the considered cover types) presented an unequivocal positive influence on the occurrence of chub. Therefore, we believe that, in further microhabitat studies, Eastern Iberian chub will select deeper microhabitats if they present some elements of cover. Such an asseveration will be supported by the aforementioned ontogenetic shifts of habitat preferences towards deeper microhabitats (Martelo et al., 2014; Martínez-Capel et al., 2009; Santos and Ferreira, 2008) and the generalised use of cover elements observed in other species of the genus (Martelo et al., 2014; Pander and Geist, 2010; Santos et al., 2004). Thereby, based on the experience gained in previous studies, where the species is either certainly present (R Muñoz-Mas et al., 2016d; Costa et al., 2012) or suspected to be present (Muñoz-Mas et al., 2017; Vezza et al., 2015), and compared to other Mediterranean species of chub, such as the Peloponnesian *S. keadicus* (Vardakas et al., 2017) or the native Iberian *S. pyrenaicus* (Martínez-Capel et al., 2009), we consider that the Eastern Iberian chub is apparently one of the *Squalius* species least prone to venture into mid-channel microhabitats. In accordance, despite the fact that distance to shore was not measured, which is a common variable in studies performed at the microhabitat scale (e.g., Vardakas et al., 2017; Martínez-Capel et al., 2009), we consider the Eastern Iberian chub to be the *Squalius* species that is most likely to remain near banks the majority of the time. Moreover, based on studies on other species of chub (Watkins et al., 1997), this behaviour could in turn be caused either by the lack of cover typical of mid-channel microhabitats or by the maximum body size achieved by the species, which is inferior to that of other species. Nevertheless, these asseverations will require confirmation, since there is evidence that, on the one hand, shoal and individual sizes (Martelo et al., 2013) and, on the other hand, season (i.e., temperature and illumination), affect *Squalius* activity (Santos and Ferreira, 2008; Baras and Nindaba, 1999).

Finally, although the species may be claimed to present a certain tendency towards limnophilia, from the microscale point of view, our results corroborate the eurytopic nature of what we suspect were Eastern Iberian chubs (Vezza et al., 2015). Consequently, although the spread of the quantiles indicated the presence of remarkable interactions between the four input variables, which could better be scrutinised with modelling approaches more transparent, for instance, fuzzy logic or generalized additive models (see e.g., Muñoz-Mas et al., 2016d, 2017), this study sheds novel insights on the habitat requirements of the species.

Therefore, we consider it will contribute to enhance environmental flow assessment and the adequate implementation of management actions focused on habitat restoration and species conservation (Martelo et al., 2013; Martínez-Capel et al., 2009; Santos et al., 2004).

5 Conclusions

This study compared four PNNs and a SVM for assessing habitat suitability of the Eastern Iberian chub. Whereas heteroscedastic and enhanced PNNs achieved the highest accuracy, these models exhibited ecologically unreliable partial dependence plots. In contrast, homoscedastic and cluster PNNs rendered ecologically reliable partial dependence plots. This could be explained by the inherent trade-off between model performance and interpretability of partial dependence plots. Based on the results of cluster PNNs, we would advocate combinations of approaches (e.g., cluster & heteroscedastic or cluster & enhanced PNNs) to balance the accuracy-interpretability trade-off. From the partial dependence plots, the Eastern Iberian chub proved to be a eurytopic species as it preferentially occurred, and hence presented the largest probability of presence, in microhabitats with cover present, low flow velocity (approx. 0.3 m/s) and intermediate depth (approx. 0.6 m) while the optimal substrate corresponded to fine gravel (64–256 mm). This ecological information on the Eastern Iberian chub should help the adequate implementation of management and restoration actions for this vulnerable species. Although several aspects require further research, we expect this study, and the annexed code, to promote the use of PNNs among scientists in general, and among ecologists and conservationists in particular for species distribution modelling and habitat suitability assessment.

6 Acknowledgments

The study has been partially funded by the national Research project IMPADAPT (CGL2013-48424-C2-1-R) with MINECO (Spanish Ministry of Economy) and Feder funds and by the Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture and Fisheries, Food and Environment). This study was also supported in part by the University Research Administration Center of the Tokyo University of Agriculture and Technology. Thanks to María José Felipe for the review on the mathematical notation and to the two anonymous reviewers who helped to improve the manuscript.

7 References

Abdollahnejad, A., Panagiotidis, D., Joybari, S.S., Surovỳ, P., 2017. Prediction of dominant forest tree species using quickbird and environmental data. Forests 8 (2). 10.3390/f8020042

Adeli, H., Panakkat, A., 2009. A probabilistic neural network for earthquake magnitude prediction. Neural Networks 22 (7), 1018–1024. 10.1016/j.neunet.2009.05.003

Ahmadlou, M., Adeli, H., 2010. Enhanced probabilistic neural network with local decision circles: A robust classifier. Integr. Comput. Aided. Eng. 17 (3), 197–210. 10.3233/ICA-

2010-0345

Alcaraz-Hernández, J.D., Martínez-Capel, F., Olaya-Marín, E.J., 2015. Length–weight relationships of two endemic fish species in the Júcar River Basin, Iberian Peninsula. J. Appl. Ichthyol. 31 (1), 246–247. 10.1111/jai.12625

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43 (6), 1223–1232. 10.1111/j.1365-2664.2006.01214.x

Ardia, D., Boudt, K., Carl, P., Mullen, K.M., Peterson, B.G., 2011. Differential evolution with deoptim. R J. 3 (1), 27–34.

Austin, M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Modell. 200 (1–2), 1–19. 10.1016/j.ecolmodel.2006.07.005

Baras, E., Nindaba, J., 1999. Seasonal and diel utilisation of inshore microhabitats by larvae and juveniles of Leuciscus cephalus and Leuciscus leuciscus. Environ. Biol. Fishes 56 (1–2), 183–197. 10.1023/A:1007594932734

Belousov, A.I., Verzakov, S.A., von Frese, J., 2002. A flexible classification approach with optimal generalisation performance: support vector machines. Chemom. Intell. Lab. Syst. 64 (1), 15–25. 10.1016/S0169-7439(02)00046-1

Ben-Hur, A., Weston, J., 2010. A User’s Guide to Support Vector Machines, in: Carugo, O., Eisenhaber, F. (Eds.), Data Mining Techniques for the Life Sciences. Totowa, NJ, pp. 223–239.

Bennetsen, E., Gobeyn, S., Goethals, P.L.M., 2016. Species distribution models grounded in ecological theory for decision support in river management. Ecol. Modell. 325, 1–12. 10.1016/j.ecolmodel.2015.12.016

Berrar, D.P., Downes, C.S., Dubitzky, W., 2003. Multiclass cancer classification using gene expression profiling and probabilistic neural networks, in: Pacific Symposium on Biocomputing. School of Biomedical Sciences, University of Ulster at Coleraine, BT521SA, Northern Ireland., pp. 5–16.

Berthold, M.R., Diamond, J., 1998. Constructive training of probabilistic neural networks. Neurocomputing 19 (1–3), 167–183. 10.1016/s0925-2312(97)00063-5

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. New York, NY (USA).

Boavida, I., Dias, V., Ferreira, M.T., Santos, J.M., 2014. Univariate functions versus fuzzy logic: Implications for fish habitat modeling. Ecol. Eng. 71, 533–538. 10.1016/j.ecoleng.2014.07.073

Bovee, K.D., 1986. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology, Instream Flow Information Paper 21. United States Fish and Wildlife Service, Biological Report.

Bovee, K.D., Lamb, B.L., Bartholow, J.M., Stalnaker, C.B., Taylor, J., Henriksen, J., 1998. Stream habitat analysis using the instream flow incremental methodology. Fort Collins, CO (USA).

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, CRC Texts in Statistical Science. New York, NY (USA).

Burrascano, P., 1991. Learning vector quantization for the probabilistic neural network. IEEE Trans. neural networks 2 (4), 458–461. 10.1109/72.88165

Casas-Mulet, R., King, E., Hoogeveen, D., Duong, L., Lakhanpal, G., Baldwin, T., et al., 2016. Two decades of ecohydraulics: trends of an emerging interdiscipline. J. Ecohydraulics. 10.1080/24705357.2016.1251296

Chang, R.K.Y., Loo, C.K., Rao, M.V.C., 2009. Enhanced probabilistic neural network with data imputation capabilities for machine-fault classification. Neural Comput. Appl. 18 (7), 791–800. 10.1007/s00521-008-0215-1

Chang, R.K.Y., Loo, C.K., Rao, M.V.C., 2008. A global k-means approach for autonomous cluster initialization of probabilistic neural network. Informatica 32 (2), 219–225.

Chasset, P.O., 2013. PNN: Probabilistic neural network for the statistical language R.

Chawla, N. V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. 10.1613/jair.953

Chen, G., Han, T.X., He, Z., Kays, R., Forrester, T., 2014. Deep convolutional neural network based species recognition for wild animal monitoring, in: 2014 IEEE International Conference on Image Processing, ICIP 2014. Paris (France), pp. 858–862.

Corne, S.A., Carver, S.J., Kunin, W.E., Lennon, J.J., van Hees, W.W.S., 2004. Predicting forest attributes in southeast Alaska using artificial neural networks. For. Sci. 50 (2), 259–276.

Cornwell, W.K., Schwilk, D.W., Ackerly, D.D., 2006. A trait-based test for habitat filtering: Convex hull volume. Ecology 87 (6), 1465–1471. 10.1890/0012-9658(2006)87[1465:ATTFHF]2.0.CO;2

Costa, R.M.S., Martínez-Capel, F., Muñoz-Mas, R., Alcaraz-Hernández, J.D., Garófano-Gómez, V., 2012. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (River Cabriel, Spain). River Res. Appl. 28 (6), 740–752. 10.1002/rra.1598

Crisci, C., Ghattas, B., Perera, G., 2012. A review of supervised machine learning algorithms and their applications to ecological data. Ecol. Modell. 240 (0), 113–122. 10.1016/j.ecolmodel.2012.03.001

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, D., 2011. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: e1071), TU Wien (Austria).

Doadrio, I., Carmona, J.A., 2006. Phylogenetic overview of the genus Squalius (Actinopterygii, Cyprinidae) in the Iberian Peninsula, with description of two new species. Cybium 30 (3), 199–214.

Eberenz, J., Verbesselt, J., Herold, M., Tsendbazar, N.-E., Sabatino, G., Rivolta, G., 2016. Evaluating the potential of proba-v satellite image time series for improving lc classification in semi-arid african landscapes. Remote Sens. 8 (12). 10.3390/rs8120987

Elith, J., Graham, C.H., 2009. Do they? How do they? Why do they differ? on finding reasons for differing performances of species distribution models. Ecography (Cop.). 32 (1), 66–77. 10.1111/j.1600-0587.2008.05505.x

Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D., Ramankutty, N., 2010. Anthropogenic transformation of the biomes, 1700 to 2000. Glob. Ecol. Biogeogr. 19 (5), 589–606. 10.1111/j.1466-8238.2010.00540.x

Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeling species distribution and change using Random Forests, in: Drew, C.A., Wiersma, Y.F., Huettmann, F. (Eds.), Predictive Species and Habitat Modeling in Landscape Ecology. New York, NY, pp. 139–

159.

Evans, M.E.K., Merow, C., Record, S., McMahon, S.M., Enquist, B.J., 2016. Towards Process-based Range Modeling of Many Species. Trends Ecol. Evol. 31 (11). 10.1016/j.tree.2016.08.005

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. 10.1214/aos/1013203451

Fukuda, S., De Baets, B., 2016. Data prevalence matters when assessing species' responses using data-driven species distribution models. Ecol. Inform. 32, 69–78. 10.1016/j.ecoinf.2016.01.005

Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M., 2013. Habitat prediction and knowledge extraction for spawning European grayling (Thymallus thymallus L.) using a broad range of species distribution models. Environ. Model. Softw. 47, 1–6. 10.1016/j.envsoft.2013.04.005

Fukuda, S., Tanakura, T., Hiramatsu, K., Harada, M., 2014. Assessment of spatial habitat heterogeneity by coupling data-driven habitat suitability models with a 2D hydrodynamic model in small-scale streams. Ecol. Inform. 10.1016/j.ecoinf.2014.10.003

Gibbs, M.S., Dandy, G.C., Maier, H.R., 2008. A genetic algorithm calibration method based on convergence due to genetic drift. Inf. Sci. (Ny). 178 (14), 2857–2869. 10.1016/j.ins.2008.03.012

Grim, J., Hora, J., 2010. Computational properties of probabilistic neural networks, in: Diamantaras, K., Duch, W., Iliadis, L.S. (Eds.), 20th International Conference on Artificial Neural Networks. Thessaloniki (Greece), pp. 31–40.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., et al., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16 (12), 1424–1435. 10.1111/ele.12189

Gumbel, E.J., 1939. La probabilité des hypotheses. Comptes Rendus l'Académie des Sci. 209, 645–647.

Gutierrez, P.A., Perez-Ortiz, M., Sanchez-Monedero, J., Fernandez-Navarro, F., Hervas-Martinez, C., 2016. Ordinal Regression Methods: Survey and Experimental Study. IEEE Trans. Knowl. Data Eng. 28 (1), 127–146. 10.1109/TKDE.2015.2457911

Habersack, H., Haspel, D., Kondolf, M., 2014. Large Rivers in the Anthropocene: Insights and tools for understanding climatic, land use, and reservoir influences. Water Resour. Res. 50 (5), 3641–3646. 10.1002/2013WR014731

Hajmeer, M., Basheer, I., 2002. A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. J. Microbiol. Methods 51 (2), 217–226. 10.1016/S0167-7012(02)00080-5

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models, Monographs on Statistics & Applied Probability. London, (UK).

Howley, T., Madden, M.G., 2005. The genetic kernel support vector machine: Description and evaluation. Artif. Intell. Rev. 24 (3–4), 379–395. 10.1007/s10462-005-9009-3

Huang, C.-L., Wang, C.-J., 2006. A GA-based feature selection and parameters optimization for support vector machines. Expert Syst. Appl. 31 (2), 231–240. 10.1016/j.eswa.2005.09.024

Hutchinson, G.E., 1957. Population studies – animal ecology and demography – concluding remarks, in: Cold Spring Harbor Symposia on Quantitative Biology. pp. 415–427.

IUCN, 2016. The IUCN Red List of Threatened Species.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31 (8), 651–666. 10.1016/j.patrec.2009.09.011

Jin, X., Cheu, R.L., Srinivasan, D., 2002. Development and adaptation of constructive probabilistic neural network in freeway incident detection. Transp. Res. Part C Emerg. Technol. 10 (2), 121–147. 10.1016/S0968-090X(01)00007-9

Jowett, I.G., Davey, A.J.H., 2007. A comparison of composite habitat suitability indices and generalized additive models of invertebrate abundance and fish presence-habitat availability. Trans. Am. Fish. Soc. 136 (2), 428–444. 10.1577/t06-104.1

Kaufman, L., Rousseeuw, P.J., 1987. Clustering by means of medoids, in: Dodge, Y. (Ed.), Statistical Data Analysis Based on the L1 Norm and Related Methods. Amsterdam (Netherlands), pp. 405–416.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43 (1), 59–69. 10.1007/BF00337288

Kusy, M.I., Zajdel, R., 2015. Application of Reinforcement Learning Algorithms for the Adaptive Computation of the Smoothing Parameter for Probabilistic Neural Network. IEEE Trans. Neural Networks Learn. Syst. 26 (9), 2163–2175. 10.1109/TNNLS.2014.2376703

LaDeau, S.L., Han, B.A., Rosi-Marshall, E.J., Weathers, K.C., 2016. The Next Decade of Big Data in Ecosystem Science. Ecosystems 1–10. 10.1007/s10021-016-0075-y

Li, L., Ma, G., 2008. Optimizing the Performance of Probabilistic Neural Networks Using PSO in the Task of Traffic Sign Recognition, in: Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. Berlin/Heidelberg (Germany), pp. 90–98.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 3 (2), 18–22.

Lorenz, A.W., Stoll, S., Sundermann, A., Haase, P., 2013. Do adult and YOY fish benefit from river restoration measures? Ecol. Eng. 61, Part A (0), 174–181. 10.1016/j.ecoleng.2013.09.027

Maceda-Veiga, A., 2013. Towards the conservation of freshwater fish: Iberian Rivers as an example of threats and management practices. Rev. Fish Biol. Fish. 23 (1), 1–22. 10.1007/s11160-012-9275-5

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2016. cluster: Cluster Analysis Basics and Extensions.

Martelo, J., Grossman, G.D., Filomena Magalhães, M., 2013. Extrinsic and intrinsic factors influence daily activity of a Mediterranean cyprinid. Ecol. Freshw. Fish 22 (2), 307–316. 10.1111/eff.12027

Martelo, J., Grossman, G.D., Porto, M., Filomena Magalhães, M., 2014. Habitat patchiness affects distribution and microhabitat use of endangered Mira chub Squalius torgalensis (Actinopterygii, Cypriniformes). Hydrobiologia 732 (1), 93–109. 10.1007/s10750-014-1850-4

Martínez-Capel, F., García De Jalón, D., Werenitzky, D., Baeza, D., Rodilla-Alamá, M., 2009. Microhabitat use by three endemic Iberian cyprinids in Mediterranean rivers (Tagus

River Basin, Spain). Fish. Manag. Ecol. 16 (1), 52–60. 10.1111/j.1365-2400.2008.00645.x

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133. 10.1007/BF02478259

Miguez, R., Georgiopoulos, M., Kaylani, A., 2010. G-PNN: A genetically engineered probabilistic neural network. Nonlinear Anal. 73 (6), 1783–1791. 10.1016/j.na.2010.04.080

Modaresi, F., Araghinejad, S., 2014. A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. Water Resour. Manag. 28 (12), 4095–4111. 10.1007/s11269-014-0730-z

Moguerza, J.M., Muñoz, A., 2006. Support vector machines with applications. Stat. Sci. 21 (3), 322–336. 10.1214/088342306000000493

Mouton, A.M., Alcaraz-Hernández, J.D., De Baets, B., Goethals, P.L.M., Martínez-Capel, F., 2011. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Model. Softw. 26 (5), 615–622. 10.1016/j.envsoft.2010.12.001

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2010. Ecological relevance of performance criteria for species distribution models. Ecol. Modell. 221 (16), 1995–2002. 10.1016/j.ecolmodel.2010.04.017

Mouton, A.M., Jowett, I., Goethals, P.L.M., De Baets, B., 2009. Prevalence-adjusted optimisation of fuzzy habitat suitability models for aquatic invertebrate and fish species in New Zealand. Ecol. Inform. 4 (4), 215–225. 10.1016/j.ecoinf.2009.07.006

Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2011. DEoptim: An R package for global optimization by differential evolution. J. Stat. Softw. 40 (6), 1–26. 10.18637/jss.v040.i06

Muniz, A.M.S., Liu, H., Lyons, K.E., Pahwa, R., Liu, W., Nobre, F.F., et al., 2010. Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait. J. Biomech. 43 (4), 720–6. 10.1016/j.jbiomech.2009.10.018

Muñoz-Mas, R., Costa, R.M.S., Martínez-Capel, F., Alcaraz-Hernández, J.D., 2017. Microhabitat competition between Iberian fish species and the endangered Júcar nase (Parachondrostoma arrigonis; Steindachner, 1866). J. Ecohydraulics 0 (0), 1–23. 10.1080/24705357.2016.1276417

Muñoz-Mas, R., Fukuda, S., Vezza, P., Martínez-Capel, F., 2016. Comparing four methods for decision-tree induction: A case study on the invasive Iberian gudgeon (Gobio lozanoi; Doadrio and Madeira, 2004). Ecol. Inform. 34, 22–34. 10.1016/j.ecoinf.2016.04.011

Muñoz-Mas, R., Lopez-Nicolas, A., Martínez-Capel, F., Pulido-Velazquez, M., 2016a. Shifts in the suitable habitat available for brown trout (Salmo trutta L.) under short-term climate change scenarios. Sci. Total Environ. 544, 686–700. 10.1016/j.scitotenv.2015.11.147

Muñoz-Mas, R., Martínez-Capel, F., Alcaraz-Hernández, J.D., Mouton, A.M., 2016b. On species distribution modelling, spatial scales and environmental flow assessment with Multi–Layer Perceptron Ensembles: A case study on the redfin barbel (Barbus haasi; Mertens, 1925). Limnol. - Ecol. Manag. Inl. Waters. 10.1016/j.limno.2016.09.004

Muñoz-Mas, R., Martínez-Capel, F., Garófano-Gómez, V., Mouton, A.M., 2014. Application of probabilistic neural networks to microhabitat suitability modelling for adult brown trout (Salmo trutta L.) in Iberian rivers. Environ. Model. Softw. 59 (0), 30–43. 10.1016/j.envsoft.2014.05.003

Muñoz-Mas, R., Martínez-Capel, F., Schneider, M., Mouton, A.M., 2012. Assessment of brown trout habitat suitability in the Jucar River Basin (Spain): Comparison of data-driven approaches with fuzzy-logic models and univariate suitability curves. Sci. Total Environ. 440, 123–131. 10.1016/j.scitotenv.2012.07.074

Muñoz-Mas, R., Papadaki, C., Martínez-Capel, F., Zogaris, S., Ntoanidis, L., Dimitriou, E., 2016c. Generalized additive and fuzzy models in environmental flow assessment: A comparison employing the West Balkan trout (Salmo farioides; Karaman, 1938). Ecol. Eng. 91, 365–377. 10.1016/j.ecoleng.2016.03.009

Muñoz-Mas, R., Vezza, P., Alcaraz-Hernández, J.D., Martínez-Capel, F., 2016d. Risk of invasion predicted with support vector machines: a case study on northern pike (Esox Lucius, L.) and bleak (Alburnus alburnus, L.). Ecol. Modell. 342, 123–134. 10.1016/j.ecolmodel.2016.10.006

Narimani, R., Narimani, A., 2013. Classification credit dataset using particle swarm optimization and probabilistic neural network models based on the dynamic decay learning algorithm. Autom. Control Intell. Syst. 1 (5), 103–112. 10.11648/j.acis.20130105.12

Öğüt, H., Aktaş, R., Alp, A., Doğanay, M.M., 2009. Prediction of financial information manipulation by using support vector machine and probabilistic neural network. Expert Syst. Appl. 36 (3), 5419–5423. 10.1016/j.eswa.2008.06.055

Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83 (2), 171–193. 10.1086/587826

Oliva, D., Cuevas, E., 2017. An introduction to machine learning, in: Advances and Applications of Optimised Algorithms in Image Processing. Intelligent Systems Reference Library, Vol 117. Cham (Switzerland), pp. 1–11.

Osuna, E., Freund, R., Girosi, F., 1997. Training support vector machines: An application to face detection, in: Anonimous (Ed.), Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Cambridge, MA (USA), pp. 130–136.

Pander, J., Geist, J., 2010. Seasonal and spatial bank habitat use by fish in highly altered rivers - a comparison of four different restoration measures. Ecol. Freshw. Fish 19 (1), 127–138. 10.1111/j.1600-0633.2009.00397.x

Parzen, E., 1962. On Estimation of a Probability Density Function and Mode. Ann. Math. Stat. 33 (3), 13–1065.

Perea, S., Doadrio, I., 2015. Phylogeography, historical demography and habitat suitability modelling of freshwater fishes inhabiting seasonally fluctuating mediterranean river systems: A case study using the iberian cyprinid Squalius valentinus. Mol. Ecol. 24 (14). 10.1111/mec.13274

Person, E., Bieri, M., Peter, A., Schleiss, A.J., 2014. Mitigation measures for fish habitat improvement in Alpine rivers affected by hydropower operations. Ecohydrology 7 (2), 580–599. http://dx.doi.org/10.1002/eco.1380

Platt, J., 2000. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: Smola, A.J., Bartlett, P.J. (Eds.), Advances in Large Margin Classifiers. Cambridge, MA (USA), pp. 61–74.

Platts, P.J., McClean, C.J., Lovett, J.C., Marchant, R., 2008. Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. Ecol. Modell. 218 (1–2), 121–134. 10.1016/j.ecolmodel.2008.06.028

Poff, N.L., Richter, B.D., Arthington, A.H., Bunn, S.E., Naiman, R.J., Kendy, E., et al., 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshw. Biol. 55 (1), 147–170. 10.1111/j.1365-2427.2009.02204.x

Quinlan, J.R., 1992. Learning with continuous classes, in: The 5th Australian Joint Conference on AI. Sydney (Australia), pp. 343–348.

R Core Team, 2015. R: A language and environment for statistical computing.

Raleigh, R.F., Zuckerman, L.D., Nelson, P.C., 1986. Habitat suitability index models and instream flow suitability curves: brown trout, revised. Washington DC, (USA).

Reineking, B., Schröder, B., 2006. Constrain to perform: Regularization of habitat models. Ecol. Modell. 193 (3–4), 675–690. 10.1016/j.ecolmodel.2005.10.003

Romão, F., Quintella, B.R., Pereira, T.J., Almeida, P.R., 2012. Swimming performance of two Iberian cyprinids: The Tagus nase Pseudochondrostoma polylepis (Steindachner, 1864) and the bordallo Squalius carolitertii (Doadrio, 1988). J. Appl. Ichthyol. 28 (1), 26–30. 10.1111/j.1439-0426.2011.01882.x

Rosenblatt, M., 1956. Remarks on Some Nonparametric Estimates of a Density Function. Ann. Math. Stat. 27 (3), 832–837. 10.1214/aoms/1177728190

Rüger, N., Schlüter, M., Matthies, M., 2005. A fuzzy habitat suitability index for Populus euphratica in the Northern Amudarya delta (Uzbekistan). Ecol. Modell. 184 (2–4), 313–328. 10.1016/j.ecolmodel.2004.10.010

Santos, J.M., Ferreira, M.T., 2008. Microhabitat use by endangered Iberian cyprinids nase Iberochondrostoma almacai and chub Squalius aradensis. Aquat. Sci. 70 (3), 272–281. 10.1007/s00027-008-8037-x

Santos, J.M., Godinho, F.N., Ferreira, M.T., 2004. Microhabitat use by Iberian nase Chondrostoma polylepis and Iberian chub Squalius carolitertii in three small streams, north-west Portugal. Ecol. Freshw. Fish 13 (3), 223–230. 10.1111/j.1600-0633.2004.00054.x

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. 10.1016/j.neunet.2014.09.003

Shin, S.J., Wu, Y., Zhang, H.H., 2014. Two-Dimensional Solution Surface for Weighted Support Vector Machines. J. Comput. Graph. Stat. 23 (2), 383–402. 10.1080/10618600.2012.761139

Shiroyama, R., Yoshimura, C., 2016. Assessing bluegill (Lepomis macrochirus) habitat suitability using partial dependence function combined with classification approaches. Ecol. Inform. 35, 9–18. 10.1016/j.ecoinf.2016.06.005

Siira, A., Erkinaro, J., Jounela, P., Suuronen, P., 2009. Run timing and migration routes of returning Atlantic salmon in the Northern Baltic Sea: implications for fisheries management. Fish. Manag. Ecol. 16 (3), 177–190. 10.1111/j.1365-2400.2009.00654.x

Sousa, R., Yevseyeva, I., da Costa, J.F.P., Cardoso, J.S., 2013. Multicriteria Models for

Learning Ordinal Data: A Literature Review, in: Yang, X.-S. (Ed.), Artificial Intelligence, Evolutionary Computing and Metaheuristics. Berlin/Heidelberg (Germany), pp. 109–138.

Specht, D.F., 1992. Enhancements to probabilistic neural networks, in: [Proceedings 1992] IJCNN International Joint Conference on Neural Networks. pp. 761–768.

Specht, D.F., 1990. Probabilistic neural networks. Neural Networks 3 (1), 109–118. 10.1016/0893-6080(90)90049-Q

Specht, D.F., 1989. Probabilistic neural networks (a one-pass learning method) and potential applications, in: Western Electronic Show and Convention (WESCON). San Francisco, CA (USA), pp. 780–785.

Specht, D.F., Romsdahl, H., 1994. Experience with adaptive probabilistic neural networks and adaptive general regression neural networks, in: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94). Orlando, FL (USA), pp. 1203–1208.

Stein, G., Chen, B., Wu, A.S., Hua, K.A., 2005. Decision tree classifier for network intrusion detection with GA-based feature selection, in: Proceedings of the 43rd Annual Southeast Regional Conference on - ACM-SE 43. New York, NY (USA), p. 136.

Storn, R., Price, K., 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. J. Glob. Optim. 11 (4), 341–359. 10.1023/A:1008202821328

Tharme, R.E., 2003. A global perspective on environmental flow assessment: Emerging trends in the development and application of environmental flow methodologies for rivers. River Res. Appl. 19 (5–6), 397–441. 10.1002/rra.736

Tierno de Figueroa, J.M., López-Rodríguez, M.J., Fenoglio, S., Sánchez-Castillo, P., Fochetti, R., 2013. Freshwater biodiversity in the rivers of the Mediterranean Basin. Hydrobiologia 719 (1), 137–186. 10.1007/s10750-012-1281-z

Tricarico, E., 2012. A review on pathways and drivers of use regarding non-native freshwater fish introductions in the Mediterranean region. Fish. Manag. Ecol. 19 (2), 133–141. 10.1111/j.1365-2400.2011.00834.x

Van Der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9.

Vapnik, V., 1995. The nature of statistical learning theory, Information Science and Statistics. New York, NY (USA).

Vardakas, L., Kalogianni, E., Papadaki, C., Vavalidis, T., Mentzafou, A., Koutsoubas, D., et al., 2017. Defining critical habitat conditions for the conservation of three endemic and endangered cyprinids in a Mediterranean intermittent river before the onset of drought. Aquat. Conserv. Mar. Freshw. Ecosyst. na (na), 1–11. 10.1002/aqc.2735

Verma, H.K., 2008. Radial basis probabilistic neural network for differential protection of power transformer. IET Gener. Transm. Distrib. 2 (1), 43–52(9). 10.1049/iet-gtd:20070037

Vezza, P., Muñoz-Mas, R., Martínez-Capel, F., Mouton, A., 2015. Random forests to evaluate biotic interactions in fish distribution models. Environ. Model. Softw. 67, 173–183. 10.1016/j.envsoft.2015.01.005

Wang, J., Neskovic, P., Cooper, L.N., 2007. Selecting Data for Fast Support Vector Machines Training, in: Kacprzyk, J. (Ed.), Trends in Neural Computation. Berlin/Heidelberg

(Germany), pp. 61–84.

Wang, J., Shen, X., Liu, Y., 2008. Probability Estimation for Large-Margin Classifiers. Biometrika 95 (1), 149–167.

Watkins, M.S., Doherty, S., Copp, G.H., 1997. Microhabitat use by 0+ and older fishes in a small English chalk stream. J. Fish Biol. 50 (5), 1010–1024. 10.1006/jfbi.1996.0363

Werbos, P.J., 1982. Applications of advances in nonlinear sensitivity analysis, in: System Modeling and Optimization. Berlin/Heidelberg (Germany), pp. 762–770.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., et al., 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14 (5), 763–773. 10.1111/j.1472-4642.2008.00482.x

Wu, T.-F., Lin, C.-J., Weng, R.C., 2004. Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. 5, 975–1005.

Wu, W.-J., Lin, S.-W., Moon, W.K., 2012. Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. Comput. Med. Imaging Graph. 36 (8), 627–633. 10.1016/j.compmedimag.2012.07.004

Yang, Z.R., Chen, S., 1998. Robust maximum likelihood training of heteroscedastic probabilistic neural networks. Neural Networks 11 (4), 739–747. 10.1016/S0893-6080(98)00024-0

Yano, K., 2016. AI for taking on the challenges of an unpredictable era. Hitachi Rev. 65 (6), 92–112.

Zhang, C., Shin, S.J., Wang, J., Wu, Y., Zhang, H.H., Liu, Y., 2013. probsvm: Class probability estimation for Support Vector Machines.

Zhong, M., Coggeshall, D., Ghaneie, E., Pope, T., Rivera, M., Georgiopoulos, M., et al., 2007. Gap-based estimation: Choosing the smoothing parameters for probabilistic and general regression neural networks. Neural Comput. 19 (10), 2840–2864. 10.1162/neco.2007.19.10.2840

Zurell, D., Elith, J., Schröder, B., 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. Divers. Distrib. 18 (6), 628–634. 10.1111/j.1472-4642.2012.00887.x