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

R. Muñoz-Mas[a*], F. Martínez-Capel[a], V. Garófano-Gómez[a], A.M. Mouton[b]

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

* Corresponding author. Tel.: (+34) 962849458; fax: (+34) 962849309

E-mail addresses: pitifleiter@hotmail.com (R. Muñoz-Mas), fmcapel@dihma.upv.es (F. Martínez-Capel), virgargm@upvnet.upv.es (V. Garófano-Gómez), ans.mouton@inbo.be (A.M. Mouton)

ABSTRACT

Probabilistic Neural Networks (PNN) have been applied and tested for the first time in microhabitat suitability modelling for adult brown trout (*Salmo trutta* L.) in Iberian rivers. The impact of prevalence on the PNN performance was studied by altering the prevalence of the training datasets. The PNN were spatially explicit evaluated in an independent river. Finally, the applicability of the PNN models to assess the minimum legal environmental flow was analysed. Prevalence did not affect significantly the results. However PNN presented some limitations regarding the output range. The habitat suitability demonstrated, a positive correlation with depth and substrate size whereas velocity showed a wider suitable range. The 0.5 prevalence PNN performed better when comparing different performance criteria, but the PNN trained with the original dataset showed higher generalisation and it allowed for a better modulation of the minimum legal environmental flow. PNN showed a valuable tool in microhabitat suitability modelling.

Keywords:

Probabilistic Neural Networks, brown trout, microhabitat suitability, prevalence, spatially explicit evaluation, Mediterranean rivers

1 Introduction

Freshwater fish are considered good indicators of water quality and biotic integrity in freshwater ecosystems (Karr, 1981). Brown trout have been specifically used as an indicator of ecological status (Ayllón et al., 2012). Therefore, insight into the habitat suitability of brown trout is crucial for conservation and cost-efficient restoration of river systems, especially in areas vulnerable to global change such as the Mediterranean streams. Studies of habitat suitability have been extended to the development of models predicting the distribution and abundance of species, as well as understanding how distribution and abundance are influenced by spatio-temporal habitat heterogeneity (Lobón-Cerviá, 2007). These ecological habitat suitability models have become an essential tool in the management and conservation of freshwater fish populations (Frank et al., 2011). The continuous univariate Habitat Suitability Curves (HSCs) are a simple and common modelling approach in studies involving the instream flow assessment (Payne and Allen, 2009), and hence several researchers have developed habitat suitability models in the form of the aforementioned HSCs (Ayllón et al., 2010; Bovee, 1978; Hayes and Jowett, 1994; Raleigh et al., 1986; Vismara et al., 2001). The Weighted Usable Area (WUA) (Bovee, 1998) derived from these models could be used in the assessment of minimum flows. For instance, the recommendations included in the Spanish norm for hydrological planning (MAGRAMA, 2008). However, several authors have suggested that considering each hydraulic variable independently may be questionable because ignoring significant interactions between variables may induce a bias (Orth and Maughan, 1982). As a consequence, the multivariate approach has increased in popularity recently (De Pauw et al., 2006). Thereby, instream flow assessment methods are in a permanent evolution driven by the imperfection and inherent constrains of each modelling technique (Lamouroux et al., 1998). Several data-driven multivariate techniques have been applied in freshwater fish habitat suitability modelling, specifically at the microhabitat scale. Ranging from simple bivariate polynomial functions (Lambert and Hanson, 1989; Vismara et al., 2001) to more complex fuzzy rule base models (Jorde et al., 2001). Logistic regression has been used by some researchers to develop habitat suitability models for brown trout (Ayllón et al., 2010; Hayes and Jowett, 1994), but Generalized Additive Models (GAMs) have been also used to predict the habitat suitability and possible interactions among variables (Jowett and Davey, 2007). In the Iberian Peninsula, the fuzzy logic approach has been applied to develop models for brown trout (Mouton et al., 2011). Particularly at the microscale being an example the habitat suitability models for middle-sized brown trout developed by Muñoz-Mas et al. (2012). Artificial Neural Networks (ANNs) and specifically the Multilayer Perceptron, have also been applied to model habitat suitability for freshwater fish at the microscale (Reyjol et al., 2001), unfortunately there are no examples available specifically focused on brown trout. Probabilistic Neural Networks (PNN) (Specht, 1990) are a promising type of ANN. These were applied successfully in pattern classification in some areas related to fish (e.g. classification of sonar signals) (Moore et al., 2003). But to our knowledge, this technique has never been applied in habitat suitability modelling at the microscale before.

An important aspect in data-driven habitat suitability modelling is the prevalence, since it can have a strong effect on model performance (Fukuda, 2013; Manel et al., 2001). The decreasing trends in brown trout populations (Almodóvar et al., 2012) or the temporary absence of the species (Lütolf et al., 2006) in addition to the sampling protocols, can lead to low prevalence databases. Several approaches have been developed (Mouton et al., 2009b) or tested (Freeman et al., 2003) to check the ability of different techniques to deal with low prevalence databases. In this context, PNN are theoretically able to cope with low prevalence databases (Specht, 1990), thus suggesting its suitability in modeling unbalanced databases. Another remarkable issue in ecological modelling is the over-fitting. Some techniques are prone in a different degree, to that phenomenon. Therefore, some authors highlighted the importance of how successful evaluation (sensu Guisan and Zimmermann, 2000) would improve the reliability of the habitat suitability models.

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

2 Materials and methods

2.1 Microhabitat data collection

The target species of this study at the microscale was the adult (body length > 20 cm) brown trout (*Salmo trutta* L.). The data samplings were carried out at low-flow conditions during late spring, summer and early autumn in the period 2007-2009 in the Guadiela and Cuervo Rivers (within the Tagus River Basin; TB) and in the Jucar and Senia Rivers (within the Jucar River Basin District; JRBD) (**Error! No s'ha trobat l'origen de la referència.**).

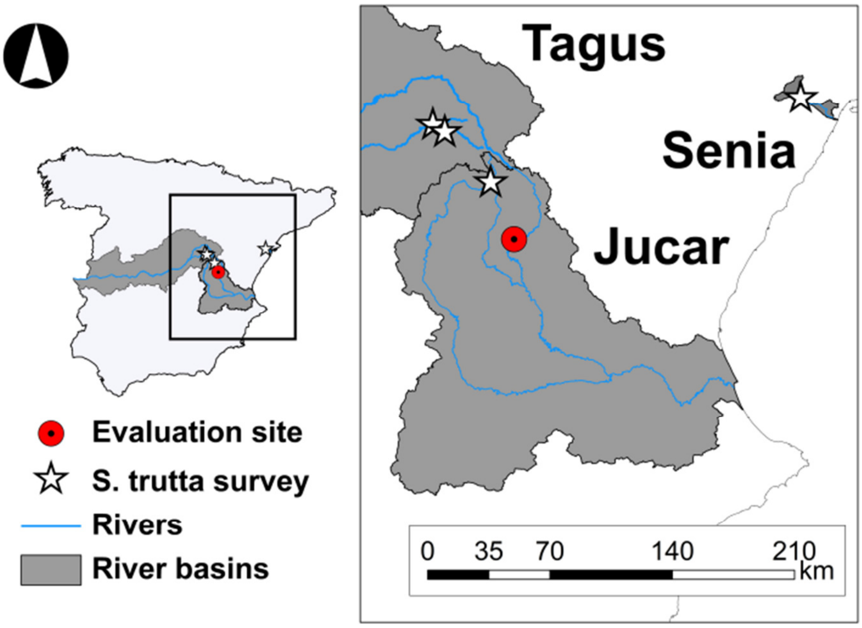

Fig. 1. In the Iberian Peninsula (left), location of the sites where microhabitat data of brown trout were collected in rivers within the Tagus River Basin and the Jucar River Basin District. Red circle shows the location where the models were evaluated in the Cabriel River.

The microhabitat study was undertaken in complete and connected HydroMorphological Units (hereafter, HMUs) and classified as pool, glide, riffle and rapid. A sort of modification of the equal effort approach was applied (Bovee, 1998) with the selection of equal areas of slow and fast water HMUs, grouping pools with glides (slow) and riffles with rapids (fast). Each HMU was studied by underwater observation (snorkelling) during daylight with minimum disturbance to the fish according to common procedures (Heggenes, 1990). This technique allows the observation of the fish behaviour thus only adult brown trout that were 'holding position' of 'feeding' were considered. Microhabitat conditions, termed as training patterns, were measured along the HMU in cross-sections classifying fish abundance into two groups as ‘absence’ (no fish observed) and ‘presence’ (at least one fish observed). The resulting sampled area per record ranged from 1.23 m^2 to 7.96 m^2. The high number of absence patterns *versus* presence patterns led to a low prevalence (average prevalence being 0.06) that ranged from 0.02 to 0.11 depending on the river (**Error! No s'ha trobat l'origen de la referència.**).

River	Year	N Presence	N Absence	Prevalence	Area surveyed (m2)
Jucar (JRBD)	2007	7	339	0.02	910
Guadiela (TRB)	2009	51	411	0.11	3189
Senia (JRBD)	2007	11	346	0.03	922
Cuervo (TRB)	2009	29	361	0.07	3065

Table 1. Sample sizes of the four campaigns for microhabitat data collection of the adult brown trout, in rivers within the Tagus River Basin (TRB) and Jucar River Basin District (JRBD).

Depth was measured with a wading rod to the nearest cm and the mean flow velocity of the water column (hereafter velocity) was measured with an electromagnetic current meter (Valeport®). The percentage of each substrate class was visually estimated around the sampling point or fish location. The substrate classification was simplified from the American Geophysical Union size scale: bedrock, boulders (>256), cobbles (64–256 mm), gravel (8–64 mm), fine gravel (2–8 mm), sand (62 mm–2 mm), silt (< 62 mm) and vegetated soil (i.e. substrate covered by macrophytes), similarly to a previous work in Iberian rivers (Martínez-Capel et al., 2009). Substrate composition was converted into a single value through the Substrate index (S), by summing weighted percentages of each substrate type as follows: S=0.08 · bedrock + 0.07 · boulder + 0.06 · cobble + 0.05 · gravel + 0.04 · fine gravel + 0.03 · sand (Mouton et al., 2011) (**Error! No s'ha trobat l'origen de la referència.**).

River	Year	Width (m)	Flow (m^3/s)	Strahler Order	Max. Velocity (m/s)	Max. Depth (m)	Substrate (Dominant)
Cuervo (TRB)	2009	8.8	0.27	2	1.03	1.24	Gravel
Guadiela (TRB)	2009	9.6	0.61	3	1.15	1.78	Cobble
Jucar (JRBD)	2007	8.4	1.05	2	1.15	1.01	Vegetation
Senia (JRBD)	2007	6.6	1.17	2	1.75	1.40	Boulder

Table 2. Characteristics of the selected rivers within the Tagus River Basin (TRB) and Jucar River Basin District (JRBD) where the microhabitat data was collected.

2.2 Development of the Probabilistic Neural Network

2.2.1 PNN theory

PNN are pattern classification radial-basis neural networks based on a Bayes-Parzen classifier (Specht, 1990). PNN basically compare the inputs with each of the measurements included in the training database and determine the probability of membership of that combination of inputs to each one of the categories present in the training database. To deal with differences in the intensity of the output, the weight of each pattern is inversely proportional to the number of training patterns in the corresponding category. Thus, the classification of certain conditions within a given category depends on the values of the inputs but not on the number of training patterns included in that category.

In the present classification problem where two categories were considered (i.e. adult brown trout 'presence' or 'absence'), the Bayes' theorem considers an input $x=[x_1, x_2, x_3, ..., x_p]$ which will be classified in the category 'presence' if the following inequality is fulfilled: $h_P \cdot i_P \cdot f_P(x) > h_A \cdot i_A \cdot f_A(x)$ where h_i is the *a priori* probability of occurrence, i_i is the cost associated with misclassification and f_i is the probability density function of the corresponding category. The aggregation of these three parameters defines the membership function. The Bayes' theorem favours a class that has a high density of training patterns in the vicinity of the unknown input ($f_i(x)$), or if the cost of misclassification (i_i) or prior probability (h_i) are high (Hajmeer and Basheer, 2002). The cost of misclassification (i_i) and the prior probability (h_i) allow the development of over-predictive models where false-positives are preferred, for instance in cancer diagnosis (Berrar et al., 2003). In our study, the *a priori* probability of occurrence (h_i) was considered 0.5 and no misclassification costs (i_i) were applied, thus both factors were neglected. In this case, the training samples must provide the information to estimate the underlying multivariate probability density functions ($f(x)$) (Specht, 1990), after this expression:

$$f(x) = \frac{1}{(2\pi)^{p/2} \prod_{j=1}^{p} \sigma_j n} \sum_{i=1}^{n} \exp\left[-\sum_{j=1}^{p} \frac{(x_j - X_j^n)^2}{2\sigma_j^2}\right] \quad (1)$$

where x is the input pattern to be classified and X^n is the i^{th} training pattern, $\sigma_1, \sigma_2, ..., \sigma_j$ are the smoothing parameters that represent the standard deviation around the mean of the p input variables, $x_1, x_2, ..., x_p$, in the present study p=3, corresponding to velocity, depth and substrate. The $\sigma_1, \sigma_2, ..., \sigma_j$ control the degree of influence in the vicinity of each training pattern. The n parameter corresponds to the total number of training patterns in the considered category (1457 absences and 98 presences). Finally the present study considered a single smoothing parameter, thus resulting in $\sigma_1 = \sigma_2 = \sigma_3$.

The smoothing parameter (σ) has a decisive impact on the PNN performance (**Error! No s'ha trobat l'origen de la referència.**). Therefore, its optimisation is recommended to obtain an optimal PNN (Hajmeer and Basheer, 2002). If the smoothing parameter is too small, the multivariate probability density function would be highly over-fitted to the training patterns, thus reducing the network's capacity to generalize. However, if the smoothing parameter is too large, the output value would be almost constant and proportional to the number of training patterns in the considered class ('presence' or 'absence'). Therefore, the values of the inputs would not play any role in the assessment of a given pattern (Zhong et al., 2005).

2.2.2 Network operation and optimisation

PNN architecture differs from other ANNs like the Multilayer Perceptron (MLP) (**Error! No s'ha trobat l'origen de la referència.**). In the case of presence-absence classification, the PNN calculates two multivariate probability density functions ($f(x)$) in parallel, one for each output category. The input pattern (e.g. every pixel in a hydraulic model) will be classified in the category of the output node that produces the most intense signal. The first layer (the input layer) is a distributing layer where x is the input pattern (i.e. a combination of velocity, depth and substrate), and it is connected to every node in the second layer (the hidden layer). The hidden layer has an equal number of neurons as there are training patterns (i.e. the 1555 collected patterns; *n*=98 presences, *m*=1457 absences). In the hidden layer, the 'difference' between the input pattern and each training pattern (XP_1, XP_2, ..., XP_n, and XA_1, XA_2, ..., XA_m) is calculated. The third layer executes the summation of the signals produced in the previous layer, but each category has an independent summation of signals. This means that the absence output node is connected only with the absence patterns (1457 connections) and the presence output node only with the presence patterns (98 connections), as demonstrated in **Error! No s'ha trobat l'origen de la referència.**. Once the sigma parameters are selected, the network is already prepared to assess any pattern. That is the main reason why PNN are considered a one-pass learning method because they are automatically trained by the patterns in the training database (Specht, 1989). Finally, the output of both nodes is standardized between 0 and 1, by dividing the results with the sum of both outputs in order to agree with other habitat suitability models.

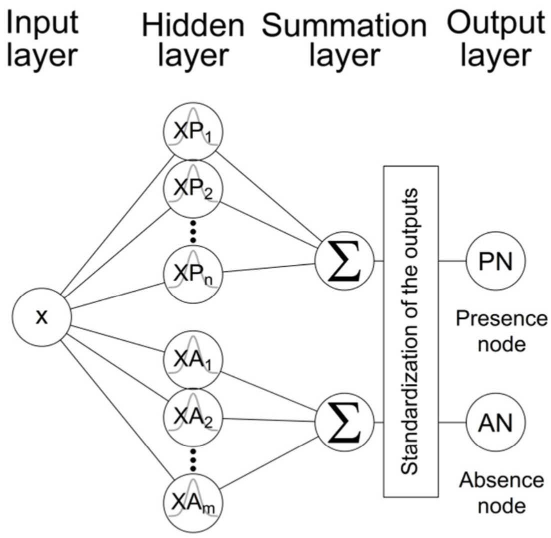

Fig. 2. General architecture of a Probabilistic Neural Network (PNN) in a presence-absence classification problem. *x* corresponds to the assessed pattern. XP and XA correspond to the presence training patterns (n=98) and the absence training patterns (m=1457) respectively.

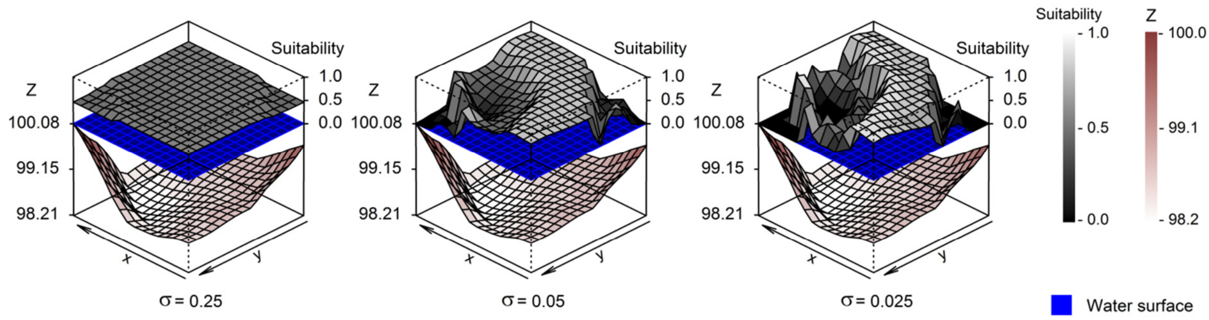

Fig. 3. Effect of the selection of different values of a single smoothing parameter (σ) in the habitat assessment of a pool at the Cabriel River (Z=elevation). The larger the σ, the smoother the

classification. Large values of the smoothing parameter sigma (0.25) do not provide the extremes of the output range (0-1) whereas lower values (0.025) provide sharper transitions achieving the extremes of the output range.

The optimisation of the PNN trained with the complete database (hereafter PNN_C) was carried out through the following steps. The three input variables (velocity, depth and substrate) were normalized and the model was developed and optimised in the R environment (R Development Core Team, 2012) by leave-one-out cross-validation. Waters (1976) introduced the use of univariate Habitat Suitability Curves (HSCs) assessing the degree of suitability of the usual microhabitat variables, such as depth or velocity, ranging from 0 and 1. Accordingly, several studies comprising difference techniques ranged the habitat suitability between 0 and 1 (Ayllón et al., 2010; Jowett and Davey, 2007; Muñoz-Mas et al., 2012). However, large values of σ typically do not provide the extreme feasible outputs (**Error! No s'ha trobat l'origen de la referència.**). Therefore, two main goals were included in the objective function, the maximization of the classification strength and the maximization of the output range. The classification strength was quantified by means of the True Skill Statistic (TSS), because this has previously been demonstrated as suitable in the modelling of unbalanced prevalence databases (Allouche et al., 2006). TSS favours a good balance between sensitivity (Sn) and specificity (Sp), while the output range was considered by subtracting the minimum output value to the maximum output value. The objective function was an aggregation of both indices. The subplex algorithm proposed by Rowan (1990) and implemented in the R environment (R Development Core Team, 2012) by King (2008) was used to optimise the smoothing parameter σ.

2.3 *Analysis of prevalence effect on PNN performance*

We analysed the effect of prevalence on PNN performance and on the output range by altering the prevalence of the training dataset but keeping the optimised σ parameter constant. Bagging was an alternative but it has been reported as extremely time-consuming (Zhong et al., 2005). Therefore, the selected datasets had to be statistically similar to the original database thus presenting similar distributions for depth, velocity and substrate. The sub-sampling methodology presented in Muñoz-Mas et al. (2012) was used to generate each of the five alternative datasets with prevalence of 0.1, 0.2, 0.3, 0.4, and 0.5. The statistical analysis, a robust generalization of Welch test (Welch, 1951) and a robust generalization of Kruskal-Wallis test (Rust and Filgner, 1984) showed no significant differences with the complete database (prevalence=0.06). These new five datasets were considered suitable for further analyses. The alternative PNN were constructed based on these five datasets and performance (TSS) and output range (minimum, maximum, quartiles, median and mean) were evaluated. The results of the leave-one-out considering the optimal σ were univariately plotted (hereafter univariate habitat suitability plots). These plots were used to check differences in the predicted suitability derived from changes in the prevalence. Subsequently, the PNN based on the 0.5 prevalence dataset (hereafter PNN_{05}) which corresponds to the ideal situation (i.e. the training dataset presents equal number of training patterns per category) was used for further analyses.

2.4 *Model transparency and ecological relevance*

Formerly, some effort has been made to improve the transparency of Neural Networks (Olden and Jackson, 2002). Following this premise, our PNN models (PNN_C and PNN_{05}) were used to assess the habitat suitability over a synthetic database that covered all possible combinations of velocity, depth and substrate. With velocity ranging from 0 to 2 m/s, depth ranged between 0 to 2 m and the substrate index ranging from 0 to 8. The entire process represented a modification of the assessment presented in Hajmeer and Basheer (2002). The assessed three-dimensional database was then plotted (hereafter multivariate habitat suitability plot) in addition to the training database assessed by means of the leave-one-out procedure. The plot divided the assessed database in slices accordingly to the substrate index illustrating brown trout habitat selection and similarities or differences on the performance derived from the consideration of different training datasets. The multivariate habitat suitability plot in combining with the univariate habitat suitability plot derived from the PNN_C was discussed in comparison with previous studies. Brown trout has been successfully introduced due to its ecological flexibility and its reputation as fine food and good sport (Klemetsen et al., 2003), therefore it presents a worldwide distribution. Accordingly it has been the target species of several studies covering a huge range of ecological conditions and spatial scales. Given the large amount of information about its habitat selection at the microscale, it was necessary to prioritize by selecting some benchmarking studies that were mainly located in the Mediterranean context and had applied multivariate techniques (Ayllón et al., 2010; Lambert and Hanson, 1989; Vismara et al., 2001). However the multivariate approach is not as widespread as the univariate. Therefore, we finally considered other studies from Europe and other continents that used multivariate approaches or due to their proven relevance (Bovee, 1978; Heggenes, 1996; Jorde et al., 2001; Jowett and Davey, 2007; Rincon and Lobon-Cervia, 1993).

2.5 *Model evaluation and transferability*

To assess the transferability of the generated PNN (PNN_C and PNN_{05}), a spatially explicit evaluation was carried out in a river reach of the Cabriel River (main tributary of the Jucar River), where an hydraulic model was available (Muñoz-Mas et al., 2012).

2.5.1 Hydraulic modelling

A 2D hydraulic simulation was done in an approximately 300 m long reach of the Cabriel River. The topographic data of the river channel and banks were collected using a Leica© Total Station. The average surface was approximately 2 m^2 per topographic point, surveying the wetted area more intensively. Substrate composition was visually estimated as aforementioned. The hydrometry was performed in 11 cross-sections, with depth and velocity measured along these sections and the resulting information was used to gauge the flow rate. Measurements were performed at three different flow rates (0.54, 1.04 and 2.75 m^3/s) and these were used for model calibration. The hydraulic modelling was carried out with River-2D© (University of Alberta, 2002) and the bed roughness was used to calibrate the model based on the depth and mean column velocity at each transect like previous studies (Jowett and Duncan, 2012). The model was considered acceptable when errors in water surface elevation were smaller than 5 cm at any cross-section, and when the patterns of the generated velocities and those measured at each cross-section were similar. The topographic data obtained in the biological evaluation survey (section **Error! No s'ha trobat l'origen de la referència.**) were used to check differences and changes in the river bed. An average difference between channel elevation in the model and control measures of 0.04±0.13 m was obtained. Therefore, the topography was considered similar and the hydraulic model acceptable for further analyses. Following this, thirty four different flows were simulated ranging from 0.05 to 6.5 m^3/s, with all of them below the bankfull stage of the river channel .

2.5.2 Biological evaluation

A new field campaign was carried out in the Cabriel River in the early summer 2012, at a flow rate of 0.89 m^3/s. Unlike the previous campaigns, the diver did not snorkel entire HMUs, but the whole area included in the aforementioned hydraulic model. The survey was carried out like the previous microhabitat surveys (Heggenes, 1990) expect velocity, depth and substrate were not recorded. Only the locations (coordinates X, Y, Z) of observed adult brown trout were recorded using a FOIF© Total Station.

2.6 PNN evaluation and applicability

The flow occurring during the biological evaluation survey was simulated with the hydraulic model. The considered PNN (PNN_C and PNN_{05}) were used to assess the habitat suitability in the entire simulated reach. The frequency analysis of the habitat suitability assessed in the trout locations and over the entire reach was compared in order to check the accuracy, the specificity and the generalization capability of the PNN.

The Spanish norm for hydrological planning (MAGRAMA, 2008) specifies the minimum legal environmental flow should be selected in accordance to the WUA-flow curves. To check the applicability of PNN for habitat assessment at different flows and for the identification of that minimum legal environmental flow, the simulated flows were assessed with the PNN_C and the PNN_{05} and the WUA-flow curves were constructed. Patterns and potential implications were discussed.

3 Results

3.1 Effects of prevalence on PNN performance

The PNN_C showed an acceptable value of the True Skill Statistic (TSS=0.35) in addition to an acceptable output range (0-0.86) (**Error! No s'ha trobat l'origen de la referència.**). The Sensitivity (Sn) was larger than the Specificity (Sp).

The univariate habitat suitability plots of flow velocity showed a suitable habitat between 0 and 1 m/s, and a maximum around 0.35 m/s (**Error! No s'ha trobat l'origen de la referència.**). The maximum velocity that was classified as 'presence' was 1.031 m/s. Depth showed two trends in the univariate habitat suitability plots. The majority of the data showed an increase of suitability as depth increased whereas a small branch showed a decrease as depth increased, this included the hull of the training patterns (**Error! No s'ha trobat l'origen de la referència.**). This effect was mainly produced by differences in the underlying substrate and is clarified in the multivariate habitat suitability plot (**Error! No s'ha trobat l'origen de la referència.**). The minimum depth classified as 'presence' was 0.16 m. The trend of the univariate habitat suitability plot for substrate was parabolic with an optimal around substrate indices corresponding to medium-to-coarse substrates, ranging from 4 to 7 (on average gravels to boulders) with a maximum at 6 (on average cobbles).

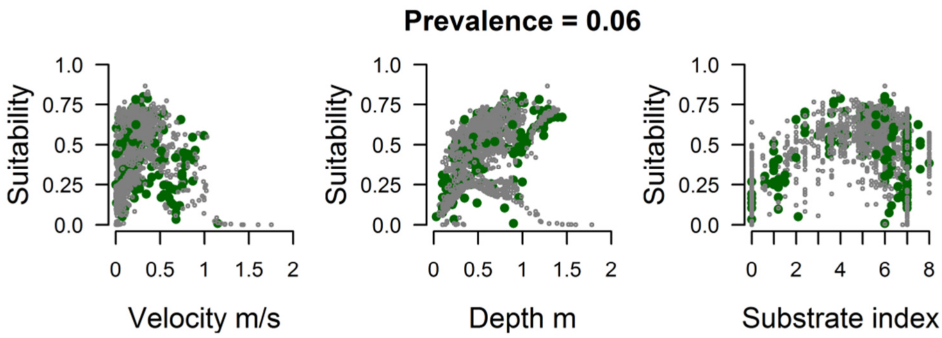

Fig. 4. Univariate habitat suitability plots based on the PNN_C. Green dots represent 'presence' data and grey dots represent 'absence' data. The patterns were classified as 'presence' when they were higher than the threshold (0.5 suitability), and as 'absence' below that threshold. The presence data were mostly correctly classified (Sn=0.77), thus they appear over the 0.5 threshold.

PNN	Prevalence	Sn	Sp	TSS	Min	1st Q	Median	Mean	3rd Q	Max
PNN_C	0.06	0.77	0.58	0.35	0	0.269	0.464	0.44	0.584	0.86
PNN_{01}	0.1	0.79	0.57	0.35	0	0.27	0.48	0.45	0.6	0.89
PNN_{02}	0.2	0.77	0.54	0.31	0	0.31	0.51	0.46	0.61	0.85
PNN_{03}	0.3	0.78	0.54	0.31	0	0.33	0.53	0.48	0.63	0.87
PNN_{04}	0.4	0.71	0.57	0.29	0	0.34	0.52	0.49	0.61	1
PNN_{05}	0.5	0.76	0.55	0.31	0	0.39	0.54	0.51	0.64	0.94

Table 3. Performance criteria (Sensitivity – Sn; Specificity – Sp; True Skill Statistic – TSS) and output statistics (minimum – Min; first quartile – 1st Q; Median; Mean; third quartile – 3rd Q; maximum – Max) corresponding to the PNN_C and the five alternative datasets and a constant smoothing parameter (σ=0.31).

Prevalence did not appear to affect the model results in a significant manner within the training datasets because the univariate habitat suitability plots generated based on the alternative PNN (**Error! No s'ha trobat l'origen de la referència.**) and the based on the PNN_C (**Error! No s'ha trobat l'origen de la referència.**) were similar. However, slight differences were present and they were observed in the multivariate plot (see **Error! No s'ha trobat l'origen de la referència.** for the PNN_C and **Error! No s'ha trobat l'origen de la referència.** for the PNN_{05}) or numerically (**Error! No s'ha trobat l'origen de la referència.**). Regarding the numerical analysis, the mean and median output values decreased with decreasing prevalence whereas the performance (TSS) slightly increased (**Error! No s'ha trobat l'origen de la referència.**). Although there were some inconsistencies in these trends, they were considered mainly due to the difficulty to extract exactly the same distribution by means of the sub-sampling procedure from a limited amount of data rather than an unclear pattern. Thus, concluding that the prevalence slightly affected the intensity of the output signals (i.e. the maximum modelled suitability increased as prevalence increased) but it did not result in a better classificatory strength (i.e. the TSS decreased as prevalence increased).

3.2 Model transparency and ecological relevance

In the multivariate habitat suitability plots based on the PNN_C, the combination of high depth and high flow velocity values resulted in the greatest habitat suitability. However, since these combinations are rare or absent in natural systems (i.e. they were not present in any training data) this was considered an anomaly due to the extrapolation of the model (**Error! No s'ha trobat l'origen de la referència.** top-right corners) and were not analysed further. The PNN_C over-predicted the 'presence' in most cases as indicated by the Sensitivity (Sn) (**Error! No s'ha trobat l'origen de la referència.**; **Error! No s'ha trobat l'origen de la referència.**). The finest substrates appeared almost unsuitable (**Error! No s'ha trobat l'origen de la referència.** Substrate index 1, 2 and 3) corresponding to the secondary branch in the univariate habitat suitability plots (**Error! No s'ha trobat l'origen de la referència.**). The habitat suitability increased broadly at medium-sized substrates but an small fringe corresponding to the shallower areas which remained unsuitable

(**Error! No s'ha trobat l'origen de la referència.** Substrate index 3, 4 and 5). At larger values of the substrate index, the aforementioned fringe increased and the suitable habitat was restricted to areas with depth larger than 0.5 m and velocity lower than 1 m/s (**Error! No s'ha trobat l'origen de la referència.** Substrate index 6, 7 and 8).

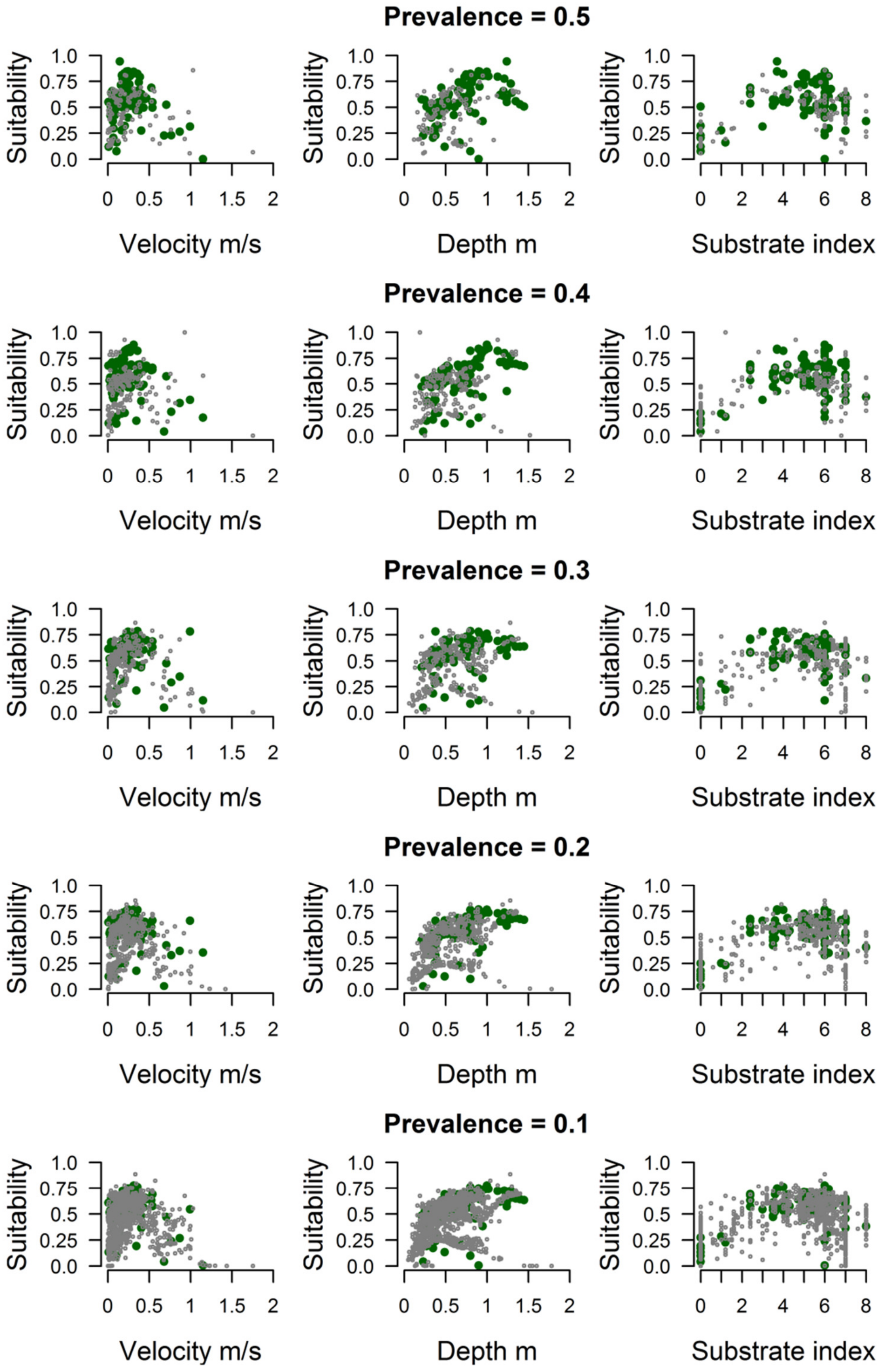

Fig. 5. Univariate habitat suitability plots based on the alternative PNN developed varying the prevalence from 0.5 to 0.1. Green dots represent 'presence' data and grey dots represent 'absence'

data. The patterns were classified as 'presence' when they were higher than the 0.5 suitability threshold, and as 'absence' if lower than that threshold.

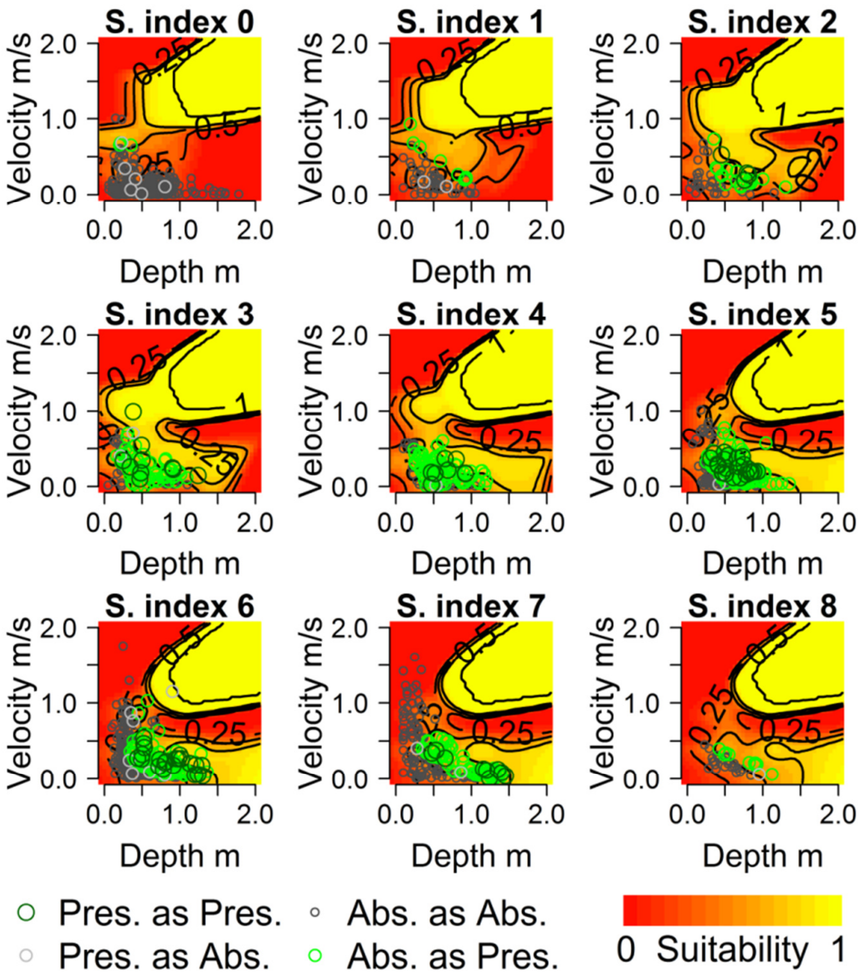

Fig. 6. Multivariate habitat suitability plot based on the PNN_C (Prevalence=0.06). Dots correspond to the training dataset with dark green for 'presence' classified as 'presence', light grey 'presence' classified as 'absence', light green 'absence' classified as 'presence', and dark grey 'absence' classified as 'absence'.

The multivariate analysis of the habitat suitability based on the PNN_{05} (**Error! No s'ha trobat l'origen de la referència.**), showed similar patterns to the PNN_C. Thus it similarly presented the anomaly in the assessment of the combination of large depth and high velocity with the maximum suitability (**Error! No s'ha trobat l'origen de la referència.** top-right corners). Obviating these areas, the PNN_{05} also showed large unsuitable areas over the finer substrates with a fringe centred on the velocity of 0.5 m/s (**Error! No s'ha trobat l'origen de la referència.** Substrate index 1, 2 and 3). The habitat suitability on average increased as the substrate index increased, achieving the maximum suitability at medium-sized substrates (**Error! No s'ha trobat l'origen de la referència.** Substrate index 4, 5 and 6). However, this PNN did not show lower habitat suitability in deeper areas. At larger values of the substrate index, the aforementioned fringe increased and the suitable habitat was also restricted to areas with depth larger than 0.5 m and velocity lower than 1 m/s (**Error! No s'ha trobat l'origen de la referència.** Substrate index 6, 7 and 8). Observing the multivariate analysis based on the PNN_{05} it showed on average a higher degree of over prediction (i.e. larger areas than in the **Error! No s'ha trobat l'origen de la referència.** were assessed with the maximum suitability, in yellow at **Error! No s'ha trobat l'origen de la referència.**) but it did not result in significant differences of the Performance criteria (**Error! No s'ha trobat l'origen de la referència.**).

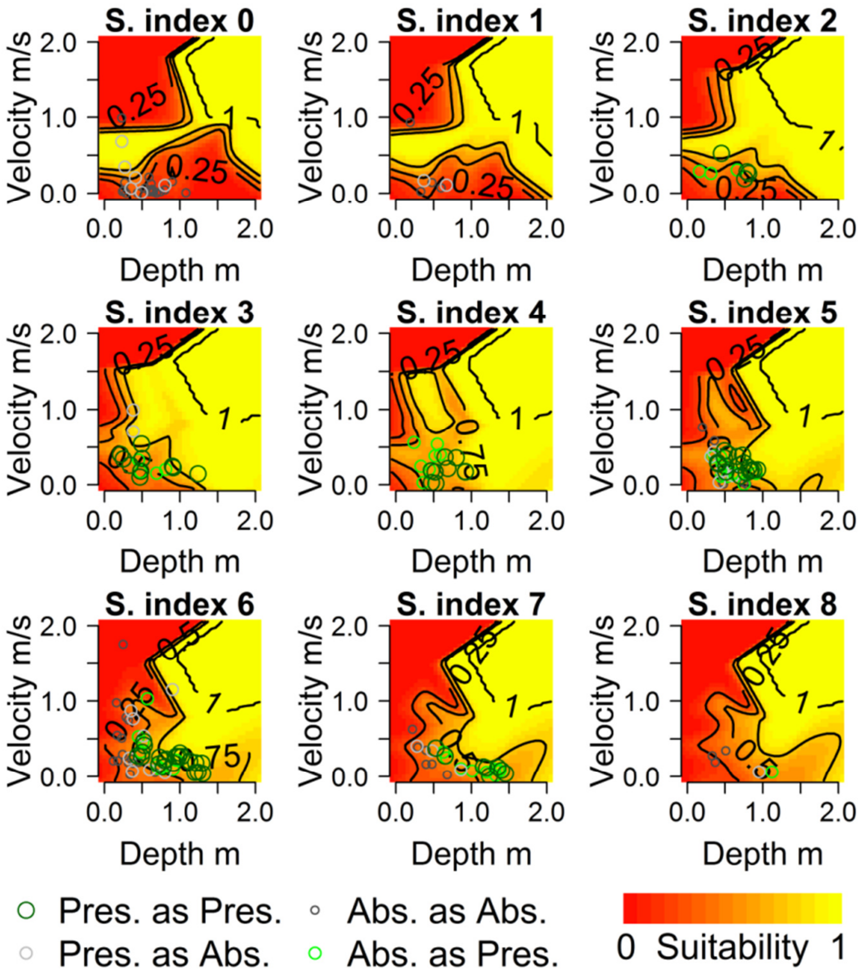

Fig. 7. Multivariate habitat suitability plot based on the PNN_{05} (Prevalence=0.5). Dots correspond to the training dataset with dark green for 'presence' classified as 'presence', light grey 'presence' classified as 'absence', light green 'absence' classified as 'presence', and dark grey 'absence' classified as 'absence'.

3.3 Model evaluation and transferability

During the biological evaluation survey, we observed thirty one adult brown trout in the reach within the hydraulic model (Q=0.89 m^3/s). The current flow was simulated providing a maximum depth of 1.4 m and a maximum velocity of 0.53 m/s, whereas the dominant and subdominant substratum were boulders (S=6; 39 %) and very fine substrate (S=0; 22 %). The PNN_C and the PNN_{05} showed similar performance, showing the higher suitability in the deeper areas (dark-green and green) and the unsuitable habitats (orange and red) in shallow areas as a consequence of the presence of low depth and fine substrate (**Error! No s'ha trobat l'origen de la referència.** Left). The PNN_C classified 65 % of the trout locations correctly, outperforming the PNN_{05} which classified correctly the 55 %. Nevertheless, both PNN provided a good trade-off between sensitivity and specificity showing crossed distributions (**Error! No s'ha trobat l'origen de la referència.** Right), although the PNN_C gave a maximum suitability of 0.8 (**Error! No s'ha trobat l'origen de la referència.** Right) and the PNN_{05} of 1 (**Error! No s'ha trobat l'origen de la referència.** Right). This confirms the aforementioned results about the effect of prevalence on model performance. On an equal footing, low prevalence slightly reduces the intensity of the signal providing lower output values.

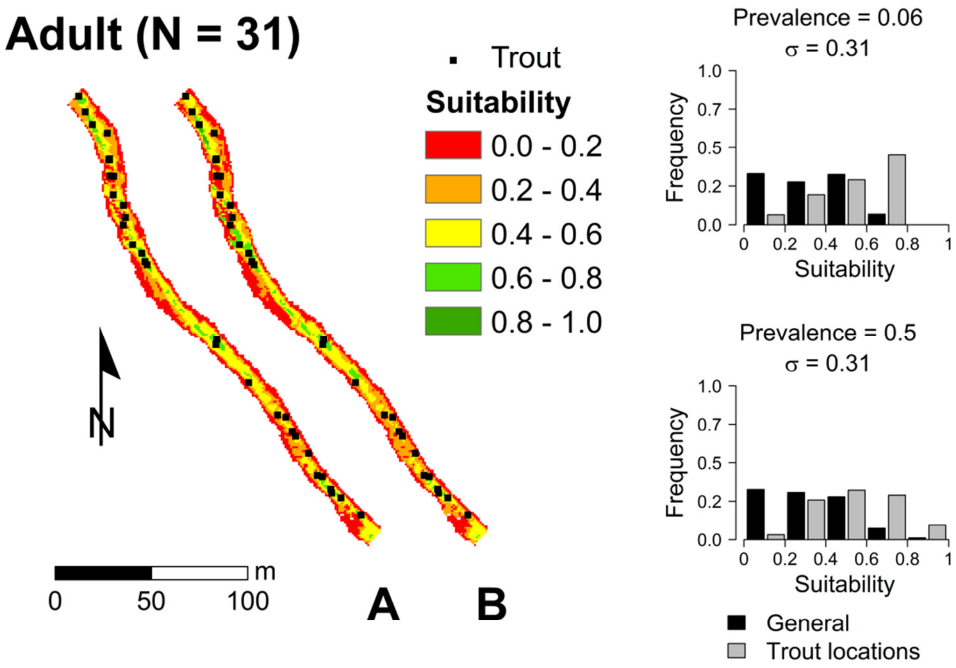

Fig. 8. On the left side, comparison of the habitat assessment on the evaluation site (Q=0.89 m^3/s) using the PNN_C (A) and the PNN_{05} (B). Red areas mean unsuitable locations and dark green areas mean locations with the maximum suitability. Black squares represent adult brown trout locations at the moment of the survey., Frequency histograms of the assessment of both PNN are on the right side. General (assessment of the entire simulated reach) is represented by black bars and Trout locations (assessment at fish locations) is represented by grey bars. It is notable that the assessment based on PNN_C did not provide the maximum suitability (range 0.8–1).

The Weighted Usable Area (WUA)-Flow curves differed depending on the considered PNN (**Error! No s'ha trobat l'origen de la referència.**). The PNN_C presented a WUA-Flow curve with an asymptotic shape (curve A). It presented an increasing trend until 3.5 m^3/s, slightly decreasing onwards but rising again for the higher simulated flows (**Error! No s'ha trobat l'origen de la referència.**). Whereas the PNN_{05} presented a monotonic increasing trend (curve B).

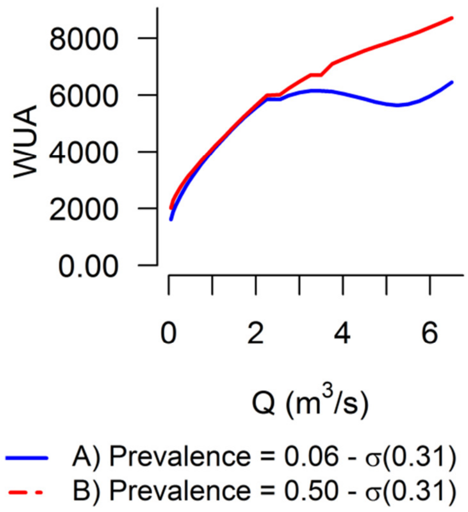

Fig. 9. Weighted Usable Area (WUA)-Flow curves for adult brown trout calculated with the 34 simulated flows ranging from 0.05 to 6.5 m^3/s. The curve A based on PNN_C presented an asymptotic shape, whereas the curve B based on PNN_{05} showed a monotonic increasing trend.

4 Discussion

4.1 Optimisation results

Mediterranean streams are characterised by strong intra and inter-annual flow variations thus becoming one of the main drivers on the observed oscillations of trout populations (Ayllón et al., 2010). This fluctuation, in combination with the sampling protocols can lead to noisy unbalanced databases including a relatively large amount of false negatives. Consequently, the selected modelling techniques might deal with this issue. The Probabilistic Neural Networks (PNN) demonstrated in this study were able to cope with these kind of databases and provided acceptable results, with an adequate True Skill Statistic (TSS) value (**Error! No s'ha trobat l'origen de la referència.**) that is in agreement with studies that have used performance criteria based on the confusion matrix and similar databases (Muñoz-Mas et al., 2012). Additionally, the PNN_C over-predicted the 'presence' providing larger values of Sensitivity (Sn) in contrast with the Specificity (Sp) (**Error! No s'ha trobat l'origen de la referència.**), which has been demonstrated to be ecologically more acceptable than under-prediction (Mouton et al., 2009b; Mouton et al., 2008).

4.2 Analysis of prevalence effect on PNN performance

The analysis of the effect of prevalence on the performance criteria (**Error! No s'ha trobat l'origen de la referència.**) in combination with the model evaluation (**Error! No s'ha trobat l'origen de la referència.**) showed that the predictive accuracy was practically unaffected (CI about 5.6 % of the mean TSS) but the maximum output value presented a decreasing trend. Some inconsistencies in these trends were observed with small fluctuations (**Error! No s'ha trobat l'origen de la referència.**). However, we considered they were mainly due to the difficulty in extracting exactly the same distribution from a limited amount of data. The procedure certainly kept the original distribution and the shapes of the univariate habitat suitability plots (**Error! No s'ha trobat l'origen de la referència.**) were fairly similar than those based on the PNN_C (**Error! No s'ha trobat l'origen de la referència.**). However, the inherent discretization in the measurement of any continuous variable (Giri and Banerjee, 2012) (i.e. velocity at the nearest mm/s, depth at the nearest cm and the substrate in the percentage of 8 classes) hindered the subsampling procedure. This method selects a real measure whilst often the most accurate choice might be between two measurements thus producing the observed differences (**Error! No s'ha trobat l'origen de la referència.**). Therefore, we considered the analysis robust enough to conclude that PNN is a suitable technique to deal with unbalanced databases, but the output values must be taken into account if values along the whole feasible range are desired. The use of an artificial database was considered because it could facilitate the subsampling procedure. However, uncertainty about the effect of prevalence in a real database remains a primary reason not to use artificial databases in this way. The results agreed with previous analyses that tested several databases and demonstrated that PNN are not constrained by the undesirable effects of unbalanced databases (Zhong et al., 2005), but the range of outputs was not considered in the aforementioned study. The range of outputs has been of major concern in the present study in order to fit the output range in common microhabitat suitability models (Jowett and Davey, 2007; Muñoz-Mas et al., 2012; Payne and Allen, 2009). Other techniques such as the fuzzy logic approach had to contend with trimmed output ranges and practical solutions have been proposed. For example, CASiMiR© (Jorde, 1997; Schneider, 2001) allow the rescaling of the outputs between 0 and 1. Nevertheless, we considered the outputs close enough to the maximum and discarded that option. The reduction on the smoothing parameter (σ) would solve this issue, thus providing values along the whole range. But it would produce overfitting because PNN are sensitive to this phenomenon (Grim and Hora, 2010; Zhong et al., 2005). The modeller should select a larger σ in order to improve the model generalization, which could lead to a smaller output range (**Error! No s'ha trobat l'origen de la referència.**). A possible solution is the consideration of the classificatory capability of PNN. Although we cannot ensure that a dichotomous output would not provide us with reliable results, the PNN will clearly be disadvantaged in comparison with other modelling techniques. Brown trout has been categorized as ecologically flexible (Klemetsen et al., 2003) and accordingly we observed a large overlapping between 'presence' and 'absence' datasets (**Error! No s'ha trobat l'origen de la referència.**). Therefore we considered that databases with better defined input-space partitions (i.e. lesser overlapping between categories) would not tend to produce trimmed outputs regardless of the prevalence in the training database. However these issues should be thoroughly explored in future research.

4.3 Model transparency and ecological relevance

The multivariate and the univariate habitat suitability plots based on the PNN_C (**Error! No s'ha trobat l'origen de la referència.** and **Error! No s'ha trobat l'origen de la referència.**) in combination with the model evaluation (**Error! No s'ha trobat l'origen de la referència.**) showed on average, a remarkable agreement with some of the most important studies. However, differences within the literature were also observed, mainly related to several factors such as differences in fish size or river types (Jowett and Davey, 2007) or in the selected sampling protocol (Heggenes et al., 1990) and their limitations. Direct observation is accurate in specific conditions and tend to underestimate the amount of individuals in shallow waters (Heggenes et al., 1990). Electrofishing is characterized by coarser resolution (Heggenes, 1996) and presents a bias related to the displacement caused by galvanotaxis (Gatz Jr et al., 1987). Neither the direct observation nor the electrofishing allow the easy observation of fish behaviour, thus databases could become noisy when including different activities (Heggenes et al., 1990). The observed differences could be due to the selected modelling approach. A recent study proposed that each modelling technique could be focused on different aspects of the training database, even predicting different habitat suitability (Fukuda et al., 2013).

Habitat selection patterns of brown trout are well established in broad terms (Ayllón et al., 2010). In near-natural rivers it has been reported to prefer relatively deep pools, occupying near-bottom locations with slow flow and medium-to-coarse substrate (Armstrong et al., 2003; Ayllón et al., 2010; Heggenes, 1996; Moyle, 2002). Accordingly, our results showed an increase of suitability as depth increases (**Error! No s'ha trobat l'origen de la referència.**) but it presented an important interaction with substrate (**Error! No s'ha trobat l'origen de la referència.**). Hence, depth did not show habitat restrictions from the substrate index of 5 onward (**Error! No s'ha trobat l'origen de la referència.**). Velocity has been reported as an important constraint on habitat suitability (Ayllón et al., 2010; Heggenes, 1996). The maximum sampled mean velocity was 1.75 m/s (Senia River), although most of the training data appeared to be below 1 m/s. The maximum velocity classified as 'presence' was 1.031 m/s (**Error! No s'ha trobat l'origen de la referència.**), therefore our results indicated a wider suitable range in comparison with previous studies (Armstrong et al., 2003) and similar in magnitude with studies conducted on large rivers (Jowett and Davey, 2007). The study did not consider water temperature, but the geographic location of the study sites suggested the presence of generally higher temperatures than previous studies. The brown trout exploits more slow flowing water in winter in comparison with summer due to differences in water temperature (Klemetsen et al., 2003). The differences in water temperature among studies, in combination with the range of differences between nose velocity and mean water column velocity (Shirvell and Dungey, 1983) support the idea that our results are ecologically significant and coherent with previous studies.

Regarding previous studies with multivariate approaches for habitat suitability modelling, the heterogeneity of results suggested that any comparison should be considered in broad terms. Vismara et al. (2001) collected data by electrofishing from an alpine river thus the degree of Mediterraneity is expected to be buffered. The developed bivariate polynomial functions based on velocity and depth showed a monotonic increment of the suitability as depth increases, whereas velocity had a weaker negative influence on it. Our results generally matched those patterns as a similar positive correlation between depth and suitability was observed (**Error! No s'ha trobat l'origen de la referència.**). Whereas velocity had a negative impact on suitability, especially over coarse substrate (**Error! No s'ha trobat l'origen de la referència.** Substrate index 6, 7 and 8). Lambert and Hanson (1989) also developed bivariate polynomial functions from data collected in small mountains streams of the King River Basin in the Sierra Nevada of California, with noticeable Mediterranean influence. The results strongly differed, since the optimal velocity and depth corresponded to 0.0 m/s and 0.5 m respectively, and both gradually tailed off as they approached their maxima, corresponding to 0.75 m/s and 2 m. It has been reported that the optimal depth for adults increases in accordance with the proportion of pools and with the maximum depth of that pools (Ayllón et al., 2010). The samplings by Lambert and Hanson (1989) were conducted in river stretches at an elevation of 1500 to 1800 m above datum, with running flows ranging from 0.7 to 0.03 m^3/s. Therefore we suggest that the observed shift of habitat suitability is partially produced by differences in habitat availability. The polynomial functions were considered rigid in the adjustment of a smooth surface to the collected data (Lambert and Hanson, 1989; Vismara et al., 2001). In contrast, the PNN was versatile in the encompassment of the suitable niche (**Error! No s'ha trobat l'origen de la referència.**).

The fuzzy logic approach has been used to develop and evaluate habitat suitability models in alpine rivers (Jorde et al., 2001). These rivers showed the maximum suitability within the velocity range of 0.3-0.9 m/s and the depth range of 0.15 - 0.5 m over medium-to-coarse substrate. Deeper areas showed higher suitability in accordance with our results. However, the morphology of braided gravel-bed rivers and the consideration of cover did not allow the proper comparison. Jowett and Davey (2007) applied Generalized Additive Models (GAMs) in modelling habitat suitability for large brown trout in a large New Zealand river (average flow > 226 m^3/s) . The partial plots showed a pointed curve for velocity with an optimal around 0.5 m/s, whereas the depth showed a wider curve with the optimum in a range between 2 and 4 m. Although body length (> 40 cm) and the river size strongly differed with our study, their results lend credibility to ours. These outcomes remarked the observed ontogenetic shift towards the selection of deeper habitats and as the availability of deep microhabitats increases (Ayllón et al., 2010). In this regard, we considered our target rivers of intermediate size in comparison with previous studies (Ayllón et al., 2010; Heggenes, 1996; Jowett and Davey, 2007; Lambert and Hanson, 1989; Rincon and Lobon-Cervia, 1993) especially concerning the available depth. Thus, the observed pattern in the multivariate habitat suitability plot (**Error! No s'ha trobat l'origen de la referència.**) (i.e. positive correlation between suitability and depth) was considered reliable. Regarding the Iberian context Ayllón *et al.* (2010) developed habitat suitability models with logistic regression in northern Iberian rivers. When considering only the rivers that better fit our range of sampling conditions (river types 4, 5 and 7) these models demonstrated a negative correlation with velocity and a positive correlation with depth, although both variables were summarized in the Froude number and the model also included cover and mesohabitat type. Therefore, proper comparison was hindered.

Regarding previous studies with univariate approaches, Ayllón et al. also developed univariate Habitat Suitability Curves (HSCs). The curves for depth were typically stable (horizontal) at the right of the optimum depth from 0.8 m onwards. Some studies on brown trout indicated such stable suitability for deep habitats (Bovee, 1978; Vismara et al., 2001). However, other studies indicated a decrease (Hayes and Jowett, 1994; Heggenes, 1996; Lambert and Hanson, 1989; Rincon and Lobon-Cervia, 1993) with optima approximately ranging from 0.5 to 1 m. This phenomenon could be a result of the modelling approach (Category II HSCs, after Bovee et al., 1998) or the absence of incoming drift (Hauer et al., 2012) rather than a negative direct effect of depth on the habitat suitability. Unfortunately, our results do not provide information about its discernment. From a univariate perspective, the mean velocity showed a wider suitable range in comparison with the optima in most of the studies in the Iberian Peninsula (0.0-0.4 m/s) (Ayllón et al., 2010; Rincon and Lobon-Cervia, 1993). However, our results were comparable with other authors (Hayes and Jowett, 1994) who did not attribute the highest suitability to very slow microhabitats, in contrast with other studies (Heggenes, 1996; Vismara et al., 2001) where habitat availability was more limited than in the present study. The substrate presented less opportunities for comparison. The model showed a maximum suitability for medium-to-coarse substrate. Generally, suitability tended to increase from fine-gravel to bedrock (**Error! No s'ha trobat l'origen de la referència.**). These results partially agreed previous studies in the Iberian peninsula (Rincon and Lobon-Cervia, 1993) and elsewhere (Bovee, 1978; Heggenes, 1996). However, these studies also showed a decrease over the bedrock whereas our model did not (**Error! No s'ha trobat l'origen de la referència.**).

Although the range of microhabitat availability was generally larger than previous European studies, the spatial distribution of the training patterns (**Error! No s'ha trobat l'origen de la referència.**) suggests that the whole distribution range of adult brown trout was not completely covered. As the model may assess extrapolated conditions unreliably (**Error! No s'ha trobat l'origen de la referència.** top-right corners), the modeller should be cautious when applying PNN outside the range of observations. In this regard PNN presented a deficiency in comparison to fuzzy logic, which allows the modification of models in areas outside of the surveyed range (Mouton et al., 2009a). To overcome this problem, further sampling campaigns should comprise extremer conditions (velocity > 1.75 m/s, depth > 1.78 m) to accurately define the suitable microniche. Fortunately these extreme conditions are rare or non-existent in the considered natural systems, allowing the application of these new models in the brown trout habitat assessment.

4.4 Model evaluation and transferability

Some authors have pointed out the difficulty to decide which models are the best, even when good model performance is achieved (Vaughan and Ormerod, 2005). Often, independent data to evaluate models is lacking, and the best model is then selected based on comparison of different performance criteria (e.g. TSS, Sn and Sp in our study). However, our results showed that evaluation based on independent data (Guisan and Thuiller, 2005) may provide valuable additional information on model performance and its generalisation capability. Specifically, the PNN_{05} performed better when comparing different performance criteria (i.e. similar TSS and larger maximum output) but the PNN_{C} showed larger generalisation capability when applied to independent data. Certainly, a reoptimised σ could improve the PNN_{05} performance or its generalisation capability. Nevertheless, once in the ideal situation (prevalence=0.5) the range of modelling techniques shall become very large and the use of PNN may become unnecessary. Aside from this, we considered that modifications of the training database would reduce the model reliability by reducing the considered variability especially when little cases are extracted. Consequently, these factors suggested that the use of the complete database should be the first option.

The transferability of habitat suitability models has been of major concern for researchers. Failures on model transferability have been reported due site-specificity and seasonal or size-related changes on habitat preferences (Fukuda, 2010). The PNN_C showed a good transferability with a 65 % of accuracy and a good specificity. This success highlighted the capability of the PNN to properly model the microhabitat suitability.

Once a single σ (σ =0.31) was selected, the Weighted Usable Area (WUA)-Flow curve presented two clear patterns depending on the prevalence of the training dataset (**Error! No s'ha trobat l'origen de la referència.**). The curve calculated with the PNN_C (curve A) presented a close-to-asymptotic shape, in comparison the curve based on the PNN_{05} (curve B) presented a monotonic increment. The Spanish norm for hydrological planning (MAGRAMA, 2008) established that the minimum legal environmental flow released to stake holders or water managers should be selected within the range of 50–80 % of the maximum WUA or considering a relevant change in the slope of the WUA-flow curve. A monotonously increasing curve akin to curve B, could only produce a single environmental flow if the break of slope was detected and not a range of minimum environmental flow. From a point of practically, the PNN_C (curve A) may be suitable for the pubic agreement about the minimum legal flow, allowing for its better modulation between the legal range (from 50 % to 80 % of the maximum WUA).

4.5 Implementation on unstudied species

Habitat suitability for the adult brown trout has been the main focus in many scientific and research projects (see aforementioned studies). The previous knowledge gained from these has allowed us to discern broadly the reliability of the developed model. In the present study we used an optimisation algorithm, but certainly with a single σ it was not obligatory. Therefore, we propose the following optimisation approach to deal with an unstudied species, thereby improving the applicability of PNN in habitat suitability modelling. The PNN optimisation should start from an arbitrary but large σ and the modeller should reduce the σ in a step-by-step procedure. The results of the leave-one-out procedure for each step should be plotted and the modeller should select an intermediate σ when a good trade-off between the bias and variance is achieved. This procedure starts from a general scope on the habitat suitability for the target species and ends with an over-fitted model lacking ecological relevance. The goal is to select an optimal model in an intermediate stage. In that sense, the PNN could be considered an intermediate step between purely data-driven models (e.g. Multilayer Perceptron) and the expert knowledge-based models such as the fuzzy rule base systems or the Category I HSCs (after Bovee, 1998).

5 Conclusions

Probabilistic Neural Networks (PNN) have been successfully applied and tested for the first time in microhabitat suitability modelling. Prevalence did not affect significantly the model results because the performance kept mostly constant but some drawbacks were observed. The PNN demonstrated the ability to deal with extremely unbalanced datasets but they presented some limitations regarding the output range. However we considered this phenomenon might be mitigated in databases with lesser overlapping between categories. Broadly, our results agreed previous studies where large brown trout has been reported to prefer relatively deep pools with slow flow and medium-to-coarse substrate (Armstrong et al., 2003; Ayllón et al., 2010; Heggenes, 1996; Moyle, 2002). The PNN trained with the 0.5 prevalence dataset, which corresponds to the ideal situation, performed better when comparing different criteria, but the PNN trained with the original dataset also presented good performance showing higher generalisation. Finally, the latter allowed for a better modulation of the minimum legal environmental flow.

A recent study focused on modelling the habitat suitability for the spawning of the European grayling (*Thymallus thymallus* L.), using a broad range of modelling techniques, (Fukuda et al., 2013) showed that Random Forest outperform any other modelling technique. Although previous studies demonstrated PNN as less competitive than other approaches (Zhong et al., 2005), results were strongly dependent on the considered databases. Therefore, once the capability of PNN to model habitat suitability is confirmed, subsequent research should focus on the comparison of PNN performance with other popular modelling techniques. Some authors suggested the need for improving models by applying abundance data because these may provide more gradual information on species' habitat selection (Fukuda et al., 2011). Indeed other models might be more appropriate when high abundances are modelled because PNN training implies the discretization of the output variable in categories. Besides the suggested comparison and following the path described in this paper the natural shift would be the use of another type of radial-basis neural networks, for instance General Regression Neural Networks (Specht, 1991). In conclusion, the present study should be considered a successful preliminary attempt in the use of PNN for habitat suitability modelling. In addition it yielded a valuable information about the PNN performance in microhabitat studies thus facilitating researchers the proper selection of the modelling technique. PNN have proven to be a useful tool in modeling habitat suitability, especially considering the use of raw databases.

Acknowledgments

The authors would like to thank the Spanish Ministry of Economy and Competitiveness for its financial support through the SCARCE project (Consolider-Ingenio 2010 CSD2009-00065). We are grateful to the colleagues who worked in the field and in the preliminary data analyses, especially Marta Bargay, Aina Hernández and David Argibay. The works were partially funded by the Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment), that also provided hydrological and environmental information about the study sites. The authors also thank the Dirección General del Agua and Infraeco for the cession of the microhabitat data. Finally, we also thank Javier Ferrer, Teodoro Estrela and Onofre Gabaldó (Confederación Hidrográfica del Júcar) for their help and the data provided. Thanks to Grieg Davies for the academic review of English.

References

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. http://dx.doi.org/10.1111/j.1365-2664.2006.01214.x.

Almodóvar, A., Nicola, G.G., Ayllón, D., Elvira, B., 2012. Global warming threatens the persistence of Mediterranean brown trout. Global Change Biol. 18(5), 1549–1560. http://dx.doi.org/10.1111/j.1365-2486.2011.02608.x.

Armstrong, J.D., Kemp, P.S., Kennedy, G.J.A., Ladle, M., Milner, N.J., 2003. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. Fish. Res. 62(2), 143-170. http://dx.doi.org/10.1016/S0165-7836(02)00160-1.

Ayllón, D., Almodóvar, A., Nicola, G.G., Elvira, B., 2010. Ontogenetic and spatial variations in brown trout habitat selection. Ecol. Freshw. Fish 19(3), 420–432. http://dx.doi.org/10.1111/j.1600-0633.2010.00426.x.

Ayllón, D., Almodóvar, A., Nicola, G.G., Parra, I., Elvira, B., 2012. A new biological indicator to assess the ecological status of Mediterranean trout type streams. Ecol. Indicators 20, 295-303. http://dx.doi.org/10.1016/j.ecolind.2012.02.028.

Berrar, D.P., Downes, C.S., Dubitzky, W., 2003. Multiclass cancer classification using gene expression profiling and probabilistic neural networks. Pac Symp Biocomput, 5-16.

Bovee, K.D., 1978. Probability of use criteria for the family Salmonidae. Instream Flow Information Paper No. 4. FWS/OBS-78/07. Fish and Wildlife Service, U.S. Department of the Interior: Washington, DC, p. 53.

Bovee, K.D., B.L. Lamb, J.M. Bartholow, C.B. Stalnaker, J. Taylor, and J. Henriksen., 1998. Stream habitat analysis using the instream flow incremental methodology. U.S. Geological Survey Information and Technology.(Report 1998-0004.), 130.

De Pauw, N., Gabriels, W., Goethals, P.L.M., 2006. River monitoring and assessment methods based on macroinvertebrates, in: Ziglio, G., Siligardi, M., Flaim, G. (Eds.), Biological monitoring of rivers: applications and perspectives. John Wiley & Sons, Ltd, pp. 111–134.

Frank, B.M., Piccolo, J.J., Baret, P.V., 2011. A review of ecological models for brown trout: towards a new demogenetic model. Ecol. Freshw. Fish 20(2), 167–198. http://dx.doi.org/10.1111/j.1600-0633.2011.00491.x.

Freeman, R.E., Stanley, E.H., Turner, M.G., 2003. Analysis and conservation implications of landscape change in the Wisconsin River floodplain, USA. Ecol. Appl. 13(2), 416–431. http://dx.doi.org/10.1890/1051-0761(2003)013[0416:AACIOL]2.0.CO;2.

Fukuda, S., 2010. Assessing transferability of genetic algorithm-optimized fuzzy habitat preference models for Japanese Medaka (Oryzias Latipes), 4th International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS): Mieres (Spain), pp. 57-62.

Fukuda, S., 2013. Effects of data prevalence on species distribution modelling using a genetic takagi-sugeno fuzzy system, Genetic and Evolutionary Fuzzy Systems (GEFS), 2013 IEEE International Workshop on, pp. 21-27.

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. Modell. Softw. 47, 1-6. http://dx.doi.org/10.1016/j.envsoft.2013.04.005.

Fukuda, S., Mouton, A.M., De Baets, B., 2011. Abundance versus presence/absence data for modelling fish habitat preference with a genetic Takagi-Sugeno fuzzy system. Environ. Monit. Assess. 184(10), 6159–6171. http://dx.doi.org/10.1007/s10661-011-2410-2.

Gatz Jr, A.J., Sale, M.J., Loar, J.M., 1987. Habitat shifts in rainbow trout: competitive influences of brown trout. Oecologia 74(1), 7–19. http://dx.doi.org/10.1007/bf00377339.

Giri, P.K., Banerjee, J., 2012. Statistical Tools and Techniques, 3rd ed. Academic Publishers, Kolkata.

Grim, J., Hora, J., 2010. Computational properties of probabilistic neural networks. 20th International Conference on Artificial Neural Networks 6354 LNCS(PART 3), 31–40. http://dx.doi.org/10.1007/978-3-642-15825-4_4.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8(9), 993–1009. http://dx.doi.org/10.1111/j.1461-0248.2005.00792.x.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135(2–3), 147–186. http://dx.doi.org/10.1016/S0304-3800(00)00354-9.

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

Hauer, C., Unfer, G., Graf, W., Leitner, P., Zeiringer, B., Habersack, H., 2012. Hydro-morphologically related variance in benthic drift and its importance for numerical habitat modelling. Hydrobiologia 683(1), 83-108. 10.1007/s10750-011-0942-7.

Hayes, J.W., Jowett, I.G., 1994. Microhabitat models of large drift-feeding brown trout in three New Zealand rivers. N. Am. J. Fish. Manage. 14(4), 710–725. http://dx.doi.org/10.1577/1548-8675(1994)014<0710:MMOLDF>2.3.CO;2.

Heggenes, J., 1990. Habitat utilization and preferences in juvenile Atlantic salmon (Salmo salar) in streams. Regul. River. 5(4), 341–354. http://dx.doi.org/10.1002/rrr.3450050406.

Heggenes, J., 1996. Habitat selection by brown trout (Salmo trutta) and young atlantic salmon (S. salar) in streams: static and dynamic hydraulic modelling. Regul. River. 12(2–3), 155–169. http://dx.doi.org/10.1002/(SICI)1099-1646(199603)12:2/3<155::AID-RRR387>3.0.CO;2-D.

Heggenes, J., Brabrand, Å., Saltveit, S., 1990. Comparison of Three Methods for Studies of Stream Habitat Use by Young Brown Trout and Atlantic Salmon. T. Am. Fish. Soc. 119(1), 101-111. 10.1577/1548-8659(1990)119<0101:cotmfs>2.3.co;2.

Jorde, K., 1997. Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken, Mitteilungen des Instituts für Wasserbau. Universität Stuttgart: Stuttgart, Germany.

Jorde, K., Schneider, M., Peter, A., Zoellner, F., 2001. Fuzzy based Models for the Evaluation of Fish Habitat Quality and Instream Flow Assessment. Proceedings of the 2001 International Symposium on Environmental Hydraulics. Tempe, Arizona.

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. T. Am. Fish. Soc. 136(2), 428-444. http://dx.doi.org/10.1577/t06-104.1.

Jowett, I.G., Duncan, M.J., 2012. Effectiveness of 1D and 2D hydraulic models for instream habitat analysis in a braided river. Ecological Engineering 48, 92–100. http://dx.doi.org/10.1016/j.ecoleng.2011.06.036.

Karr, J.R., 1981. Assessment of biotic integrity using fish communities. Fisheries 6(6), 21–27. http://dx.doi.org/10.1577/1548-8446(1981)006<0021:aobiuf>2.0.co;2.

King, A.A., 2008. Subplex optimization algorithm. Package 'subplex'. R package, version 1.1–3. http://www.cran.r-project.org/web/packages/subplex/subplex.pdf.

Klemetsen, A., Amundsen, P.A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F., Mortensen, E., 2003. Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): A review of aspects of their life histories. Ecol. Freshw. Fish 12(1), 1-59. 10.1034/j.1600-0633.2003.00010.x.

Lambert, T.R., Hanson, D.F., 1989. Development of habitat suitability criteria for trout in small streams. Regul. River. 3(1), 291-303. 10.1002/rrr.3450030128.

Lamouroux, N., Capra, H., Pouilly, M., 1998. Predicting habitat suitability for lotic fish: linking statistical hydraulic models with multivariate habitat use models. Regul. River. 14(1), 1–11. http://dx.doi.org/10.1002/(sici)1099-1646(199801/02)14:1<1::aid-rrr472>3.0.co;2-d.

Lobón-Cerviá, J., 2007. Numerical changes in stream-resident brown trout (Salmo trutta): uncovering the roles of density-dependent and density-independent factors across space and time. Can. J. Fish. Aquat. Sci. 64(10), 1429–1447. http://dx.doi.org/10.1139/f07-111.

Lütolf, M., Kienast, F., Guisan, A., 2006. The ghost of past species occurrence: Improving species distribution models for presence-only data. J. Appl. Ecol. 43(4), 802–815. http://dx.doi.org/10.1111/j.1365-2664.2006.01191.x.

MAGRAMA, 2008. Orden ARM/2656/2008, de 10 de septiembre, por la que se aprueba la instrucción de planificación hidrológica.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38(5), 921–931. http://dx.doi.org/10.1046/j.1365-2664.2001.00647.x.

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). Fisheries Manag. Ecol. 16(1), 52–60. http://dx.doi.org/10.1111/j.1365-2400.2008.00645.x.

Moore, D.R.J., Breton, R.L., MacDonald, D.B., 2003. A comparison of model performance for six quantitative structure-activity relationship packages that predict acute toxicity to fish. Environ. Toxicol. Chem. 22(8), 1799–1809. http://dx.doi.org/10.1897/00-361.

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. Modell. Softw. 26(5), 615–622. http://dx.doi.org/10.1016/j.envsoft.2010.12.001.

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2009a. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environ. Modell. Softw. 24(8), 982–993. http://dx.doi.org/10.1016/j.envsoft.2009.02.005.

Mouton, A.M., De Baets, B., Van Broekhoven, E., Goethals, P.L.M., 2009b. Prevalence-adjusted optimisation of fuzzy models for species distribution. Ecol. Model. 220(15), 1776–1786. http://dx.doi.org/10.1016/j.ecolmodel.2009.04.020.

Mouton, A.M., Schneider, M., Peter, A., Holzer, G., Müller, R., Goethals, P.L.M., De Pauw, N., 2008. Optimisation of a fuzzy physical habitat model for spawning European grayling (Thymallus thymallus L.) in the Aare river (Thun, Switzerland). Ecol. Model. 215(1–3), 122–132. http://dx.doi.org/10.1016/j.ecolmodel.2008.02.028.

Moyle, P.B., 2002. Inland Fishes of California. University of California Press, Berkeley, California.

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. http://dx.doi.org/10.1016/j.scitotenv.2012.07.074.

Olden, J.D., Jackson, D.A., 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 154(1–2), 135–150. http://dx.doi.org/10.1016/s0304-3800(02)00064-9.

Orth, D.J., Maughan, O.E., 1982. Evaluation of the incremental methodology for recommending instream flows for fishes. T. Am. Fish. Soc. 111(4), 413–445. http://dx.doi.org/10.1577/1548-8659(1982)111<413:eotimf>2.0.co;2.

Payne, T.R., Allen, M.A., 2009. Application of the use-to-availability electivity ratio for developing habitat suitability criteria in PHABSIM instream flow studies. Proceedings of the 7th International Symposium on Ecohydraulics, Concepción, Chile, January 12-16, 2009.

R Development Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria.

Raleigh, R.F., Zuckerman, L.D., Nelson, P.C., 1986. Habitat suitability index models and instream flow suitability curves: brown trout, revised. Biological Report 82 (10.124). FWS/OBS - 82/10.71. Fish and Wildlife Service. U.S. Department of Interior: Washington, DC, p. 65.

Reyjol, Y., Lim, P., Belaud, A., Lek, S., 2001. Modelling of microhabitat used by fish in natural and regulated flows in the river Garonne (France). Ecol. Model. 146(1–3), 131–142. http://dx.doi.org/10.1016/s0304-3800(01)00301-5.

Rincon, P.A., Lobon-Cervia, J., 1993. Microhabitat use by stream-resident brown trout: bioenergetic consequences. T. Am. Fish. Soc. 122(4), 575-587.

Rowan, T.H., 1990. Functional stability analysis of numerical algorithms. Doctoral Dissertation. University of Texas: Austin, Texas.

Rust, S.W., Filgner, M.A., 1984. A modification of the kruskal-wallis statistic for the generalized behrens-fisher problem. Commun. Stat. - Theory and Methods 13(16), 2013-2027. 10.1080/03610928408828810.

Schneider, M., 2001. Habitat - und Abflussmodellierung für Fließgewässer mit unscharfen Berechnungsansätzen., Mitteilungen des Instituts für Wasserbau. Universität Stuttgart: Stuttgart.

Shirvell, C.S., Dungey, R.G., 1983. Microhabitats Chosen by Brown Trout for Feeding and Spawning in Rivers. T. Am. Fish. Soc. 112(3), 355-367. 10.1577/1548-8659(1983)112<355:mcbbtf>2.0.co;2.

Specht, D.F., 1989. Probabilistic neutral networks (a one-pass learning method) and potential applications, Western Electronic Show and Convention (WESCON). Western Periodicals Conference: San Francisco, California, pp. 780-785.

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

Specht, D.F., 1991. A General regression neural network. IEEE T. Neural. Networ. 2(6), 568-576. http://dx.doi.org/10.1109/72.97934.

Vaughan, I.P., Ormerod, S.J., 2005. The continuing challenges of testing species distribution models. J. Appl. Ecol. 42(4), 720–730. http://dx.doi.org/10.1111/j.1365-2664.2005.01052.x.

Vismara, R., Azzellino, A., Bosi, R., Crosa, G., Gentili, G., 2001. Habitat suitability curves for brown trout (Salmo trutta fario L.) in the River Adda, Northern Italy: Comparing univariate and multivariate approaches. Regul. River. 17(1), 37–50. http://dx.doi.org/10.1002/1099-1646(200101/02)17:1<37::AID-RRR606>3.0.CO;2-Q.

Waters, B.F., 1976. A Methodology for Evaluating the Effects of Different Streamflows on Salmonid Habitat - Proceedings of the Symposium and Specialty Conference on Instream Flow Needs. American Fisheries Society: Bathesda, MD, p. 13.

Welch, B.L., 1951. On the comparison of several mean values: An alternative approach. Biometrika 38(3/4), 6.

Zhong, M., Hecker, J., Maidhoff, I., Shibly, P., Georgiopoulos, M., Anagnostopoulos, G., Mollaghasemi, M., 2005. Probabilistic Neural Network: Comparisons of the Cross-Validation Approach and a Fast Heuristic to choose the Smoothing Parameters, Artificial Neural Networks in Engineering: St. Louis, Missouri.