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

1. Mun˜ oz-Mas[a*,*](#_bookmark0)∗, F. Martínez-Capel[a](#_bookmark0), J.D. Alcaraz-Hernández[a](#_bookmark0), A.M. Mouton[b](#_bookmark1)

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*

a b s t r a c t

*Keywords:*

Artiﬁcial neural networks

*Barbus haasi*

Data mining

Species distribution modelling Uncertainty analysis

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

* 1. **Introduction**

Ecological models for the quantitative prediction of species dis- tributions are key to understanding the realised niche of species and its implication for species conservation in relation to global change ([Austin,](#_bookmark34) [2007).](#_bookmark34) Therefore, ecological models have increasingly received attention due to their wide management applications in the context of biogeography, conservation biology and climate change studies ([Mouton](#_bookmark65) [et al.,](#_bookmark65) [2010).](#_bookmark65) Many studies on ecologi- cal modelling have focused on explanation rather than prediction ([Elith](#_bookmark19) [and](#_bookmark19) [Leathwick,](#_bookmark19) [2009);](#_bookmark19) however, differences in the life-history or in the gene ﬂow of ﬁsh assemblages could result in different realised niches ([Mouton](#_bookmark65) [et al.,](#_bookmark65) [2010).](#_bookmark65) Abiotic factors, together with dispersal and biotic interactions, are often considered the three elements that shape the ecological niche by determining species distribution and abundance ([Barve](#_bookmark38) [et al.,](#_bookmark38) [2011).](#_bookmark38) However, ecologi- cal models have usually focused on abiotic factors only ([Boulangeat](#_bookmark42)

∗ Corresponding author. Tel.: +34 962849458; fax: +34 962849309.

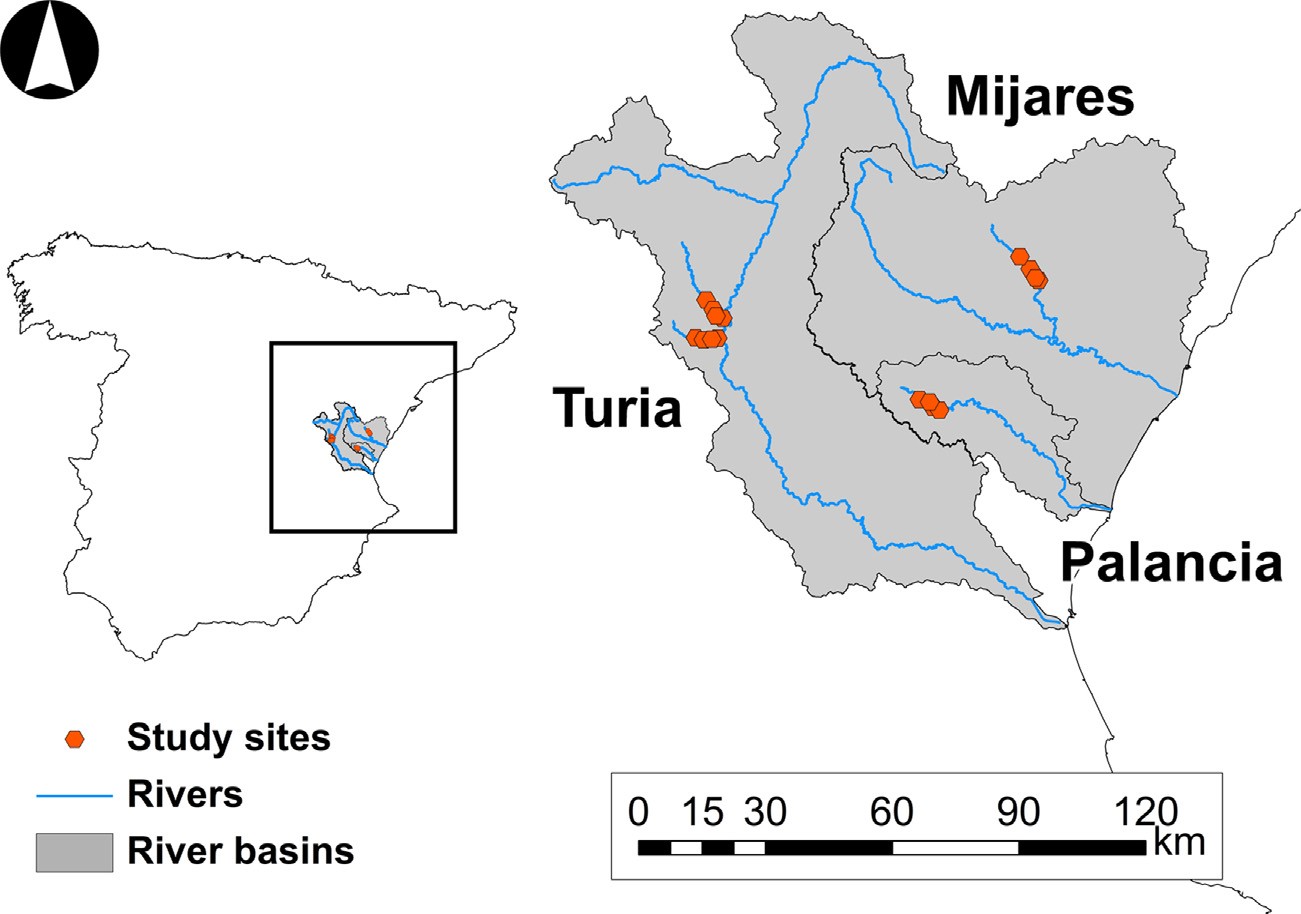
*E-mail addresses:* [pitiﬂeiter@hotmail.com](mailto:pitifleiter@hotmail.com)

(R. Mun˜ oz-Mas), [fmcapel@dihma.upv.es](mailto:fmcapel@dihma.upv.es) (F. Martínez-Capel), [jdalcaraz@gmail.com](mailto:jdalcaraz@gmail.com) (J.D. Alcaraz-Hernández), [ans.mouton@inbo.be](mailto:ans.mouton@inbo.be) (A.M. Mouton).

[et al.,](#_bookmark42) [2012),](#_bookmark42) and very few studies in freshwater ﬁsh ecology have explicitly included biotic variables ([Elith](#_bookmark19) [and](#_bookmark19) [Leathwick,](#_bookmark19) [2009)](#_bookmark19) to explore biotic interactions and consumer-resource dynamics ([Soberón,](#_bookmark63) [2007).](#_bookmark63) The consideration of these three elements (i.e. abiotic, biotic and dispersal factors) do not allow for simple statis- tical analysis because the data collected often exhibit non-linear and complex data structures ([Crisci](#_bookmark13) [et al.,](#_bookmark13) [2012).](#_bookmark13) Consequently, there is a need for new and innovative approaches to under- stand the complex structure of living systems ([Larocque](#_bookmark43) [et al.,](#_bookmark43) [2011).](#_bookmark43)

Several sophisticated modelling techniques have been applied in the ecological modelling of ﬁsh species, ranging from linear to multivariate and machine learning techniques such as Artiﬁcial Neural Networks (ANN) ([Brosse](#_bookmark45) [and](#_bookmark45) [Lek,](#_bookmark45) [2000,](#_bookmark45) [Mun˜ oz-Mas](#_bookmark67) [et al.,](#_bookmark67) [2014,](#_bookmark67) [Palialexis](#_bookmark51) [et al.,](#_bookmark51) [2011).](#_bookmark51) The most popular ANN architecture has been the Multilayer Perceptron (MLP) paradigm because it is con- sidered to be able to approximate any continuous function ([Olden](#_bookmark46) [et al.,](#_bookmark46) [2008).](#_bookmark46) Formerly, MLP was referred to as a ‘black box’ because it provided little explanatory insight into the relative inﬂuence of variables in the prediction process ([Olden](#_bookmark70) [and](#_bookmark70) [Jackson,](#_bookmark70) [2002).](#_bookmark70) To date, an enormous effort has been made to develop methods that clarify variable importance and interactions ([Gevrey](#_bookmark26) [et al.,](#_bookmark26) [2006;](#_bookmark26) [Lek](#_bookmark26) [et al.,](#_bookmark26) [1996;](#_bookmark26) [Olden](#_bookmark26) [and](#_bookmark26) [Jackson,](#_bookmark26) [2002),](#_bookmark26) and consequently, MLPs

.



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

should no longer be treated as ‘black box’ models ([Özesmi](#_bookmark49) [et al.,](#_bookmark49) [2006).](#_bookmark49)

There are several examples of single MLP applications in fresh- water ﬁsh ecology ([Park](#_bookmark54) [and](#_bookmark54) [Chon,](#_bookmark54) [2007).](#_bookmark54) For instance, MLPs have been successfully applied to model ﬁsh ecology through a broad range of ecosystems ([Brosse](#_bookmark45) [and](#_bookmark45) [Lek,](#_bookmark45) [2000;](#_bookmark45) [Gevrey](#_bookmark45) [et al.,](#_bookmark45) [2006;](#_bookmark45) [Kemp](#_bookmark45) [et al.,](#_bookmark45) [2007;](#_bookmark45) [Laffaille](#_bookmark45) [et al.,](#_bookmark45) [2003)](#_bookmark45) and in some cases out- performing other statistical approaches ([Baran](#_bookmark35) [et al.,](#_bookmark35) [1996;](#_bookmark35) [Lek](#_bookmark35) [et al.,](#_bookmark35) [1996).](#_bookmark35) Despite those successful studies, it has been demon- strated that single models (e.g. a single MLP) do not necessarily perform consistently, resulting in divergent predictions ([Buisson](#_bookmark14) [et al.,](#_bookmark14) [2010;](#_bookmark14) [Fukuda](#_bookmark14) [et al.,](#_bookmark14) [2011a,](#_bookmark14) [2013).](#_bookmark14) The use of model ensembles has been emphasised to overcome this phenomenon ([Araújo](#_bookmark31) [and](#_bookmark31) [New,](#_bookmark31) [2007).](#_bookmark31) The Multilayer Perceptron Ensemble (MLP Ensemble, [Hansen](#_bookmark32) [and](#_bookmark32) [Salamon,](#_bookmark32) [1990)](#_bookmark32) has proven to be proﬁcient in several areas of ecology ([Palialexis](#_bookmark51) [et al.,](#_bookmark51) [2011;](#_bookmark51) [Watts](#_bookmark51) [and](#_bookmark51) [Worner,](#_bookmark51) [2008),](#_bookmark51) but has rarely been applied in freshwater ecosystems ([Mun˜ oz-Mas](#_bookmark67) [et al.,](#_bookmark67) [2014).](#_bookmark67)

Fish communities in Mediterranean rivers are an interesting tar- gets to develop these novel statistical approaches ([Hopkins](#_bookmark33) [II](#_bookmark33) [and](#_bookmark33) [Burr,](#_bookmark33) [2009),](#_bookmark33) particularly communities dominated by cyprinids, as they are characterised by a high number of endemic species for which there is insufﬁcient knowledge about their ecology ([Ferreira](#_bookmark20) [et al.,](#_bookmark20) [2007).](#_bookmark20) Furthermore, endemic species tend to facilitate a more robust analysis of species–environment relationships. In this paper, we focused on the redﬁn barbel (*Barbus haasi* Mertens, 1925), a rheophilic small barbel (maximum body-length 30 cm) that is endemic to the Iberian Peninsula ([Bianco,](#_bookmark39) [1998)](#_bookmark39) and categorised as vulnerable ([Freyhof](#_bookmark21) [and](#_bookmark21) [Brooks,](#_bookmark21) [2011).](#_bookmark21) Their populations have decreased markedly, with pollution and the presence of exotic species being the main factors involved in the decline ([Perea](#_bookmark56) [et al.,](#_bookmark56) [2011).](#_bookmark56) Although redﬁn barbel has been the subject of numerous studies addressing its life-history, home-range, habitat prefer- ences and the effects of pollutants ([Aparicio](#_bookmark30) [and](#_bookmark30) [De](#_bookmark30) [Sostoa,](#_bookmark30) [1999;](#_bookmark30) [Aparicio,](#_bookmark30) [2002;](#_bookmark30) [Figuerola](#_bookmark30) [et al.,](#_bookmark30) [2012;](#_bookmark30) [Grossman](#_bookmark30) [and](#_bookmark30) [De](#_bookmark30) [Sostoa,](#_bookmark30) [1994),](#_bookmark30) a knowledge gap remains on the impact of biotic variables such as the density of accompanying ﬁsh species or invertebrate predictors in its ecological niche.

Therefore, the objective of this study was: (1) to test the pro- ﬁciency of the MLP Ensembles to model the ecological niche of freshwater ﬁsh species and (2) to test whether biotic variables affect the distribution of redﬁn barbel. To achieve these aims using MLP

Ensembles, two different models of redﬁn barbel were developed. The ﬁrst considered only physical habitat variables, the second included biotic and physical habitat variables.

* 1. **Materials and methods**
     1. *Data collection*

The study was conducted at the meso-scale in every summer, between 2003 and 2006. The study sites were located in the head- waters of the Ebron and Vallanca Rivers (Turia River tributaries), the Palancia River and the Villahermosa River (Mijares River Tributary) ([Fig. 1)](#_bookmark3) which approximately correspond to the southern limits of redﬁn barbel distribution ([Perea](#_bookmark56) [et al.,](#_bookmark56) [2011).](#_bookmark56) All the study sites were in unregulated streams and therefore a wide ﬂow range was sampled (i.e. from 0.02 m3/s to 1.84 m3/s). For complete climatic description of the study area, see [Alcaraz-Hernández](#_bookmark27) [et al.](#_bookmark27) [(2011)](#_bookmark27) and [Mouton](#_bookmark64) [et al.](#_bookmark64) [(2011).](#_bookmark64)

* + - 1. *Physical habitat survey*

The physical habitat was assessed in every 300 m reach using an adaptation of the Basinwide Visual Estimation Technique ([Dolloff](#_bookmark15) [et al.,](#_bookmark15) [1993).](#_bookmark15) The approach stratiﬁes the study site by HydroMorphological Units (hereafter called HMUs) classiﬁed as: pools, glides, rifﬂes, and rapids (see [Alcaraz-Hernández](#_bookmark27) [et al.,](#_bookmark27) [2011](#_bookmark27) for further details). Once an HMU was categorised, its phys- ical attributes were recorded. They were, length, average width, obtained from three cross-sections corresponding to ¼, ½, and ¾ of the total length, mean depth (hereafter as depth), calculated from nine points corresponding to the measurements taken at each of the aforementioned cross-sections and the maximum depth, mea- sured at the corresponding point. Percentage of shading over the channel, percentage of embeddedness, pieces of woody debris and percentage of the substrate types following a simpliﬁed classiﬁca- tion from the American Geophysical Union ([Martínez-Capel](#_bookmark55) [et al.,](#_bookmark55) [2009;](#_bookmark55) [Mun˜ oz-Mas](#_bookmark55) [et al.,](#_bookmark55) [2012)](#_bookmark55) were visually estimated and summ- arised in the substrate index ([Mouton](#_bookmark64) [et al.,](#_bookmark64) [2011).](#_bookmark64) In addition, the cover index ([García](#_bookmark24) [de](#_bookmark24) [Jalón](#_bookmark24) [and](#_bookmark24) [Schmidt,](#_bookmark24) [1995)](#_bookmark24) was determined. This index characterises the available refuge due to caves, shading, substrate, submerged vegetation and water depth by assigning six scores from 0 (no refuge) to 5 (maximum score), and the weighted aggregation of these scores produces an index range from 0 to 10.

**Table 1**

Code, summary, units and description of the variables included in the MLP Ensemble models.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable code | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | Units | Description |
| River.Reach | 1 | 4 | 8 | 8.108 | 12 | 16 | – | Study site code |
| Year | 2003 | 2004 | 2005 | 2005 | 2006 | 2006 | – | Date |
| Meso.type | 1 | 2 | 3 | 2.95 | 4 | 4 | – | Mesohabitat type |
| Meso.diversity | 0.2 | 0.62 | 0.7 | 0.68 | 0.76 | 0.99 | – | Reach mesohabitat diversity |
| Length | 8.6 | 19.1 | 24.36 | 26.92 | 31.5 | 54.7 | m | Length |
| Width | 1.26 | 3.43 | 4.79 | 4.66 | 5.83 | 8.8 | m | Width |
| Depth | 0.04 | 0.22 | 0.32 | 0.35 | 0.46 | 0.79 | m | Mean depth |
| M.Depth | 0.15 | 0.43 | 0.63 | 0.64 | 0.83 | 1.23 | m | Maximum depth |
| Velocity | 0.01 | 0.09 | 0.24 | 0.3 | 0.42 | 1.06 | m/s | Mean ﬂow velocity |
| Substrate | 2.65 | 4.9 | 5.2 | 5.22 | 5.7 | 8 | – | Substrate index |
| Embeddedness | 0 | 0 | 15 | 29.35 | 50 | 100 | % | % mud covering substrate |
| Cover | 1 | 2.75 | 3.5 | 3.67 | 4.25 | 7.5 | – | Cover index |
| Shadow | 0 | 20 | 60 | 54.95 | 85 | 100 | % | % shading |
| Wood.debris | 0 | 0 | 0 | 0.01 | 0 | 0.16 | pieces/m2 | Woody debris |
| Elevation | 605 | 655 | 743 | 745.8 | 792 | 968 | m | Reach elevation above datum |
| Slope | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.04 | m/m | Reach slope |
| D.redﬁn | 0 | 0 | 0 | 2.61 | 2.66 | 31.22 | ind./100 m2 | Density of *Barbus haasi* |
| D.b.trout | 0 | 1.32 | 4.48 | 11.19 | 16.12 | 86.47 | ind./100 m2 | Density of *Salmo trutta* |
| D.eel | 0 | 0 | 0 | 0.9 | 0 | 20.74 | ind./100 m2 | Density of *Anguilla anguilla* |
| D.r.trout | 0 | 0 | 0 | 3.64 | 2.18 | 42.58 | ind./100 m2 | Density of *Oncorhynchus mykiss* |
| D.cyprinids | 0 | 0 | 0 | 12.31 | 11.09 | 198.3 | ind./100 m2 | Cyprinids density |
| Inv.density | 0 | 1930 | 4680 | 7910 | 9590 | 56,010 | ind./m2 | Invertebrates density |
| Inv.richness | 0 | 16 | 19 | 18.55 | 22 | 34 | – | Invertebrates richness |
| Inv.diversity | 0 | 0.34 | 0.41 | 0.39 | 0.46 | 0.56 | – | Invertebrates diversity |
| Inv.biomass | 0 | 0 | 0.07 | 0.39 | 0.27 | 6.02 | g/m2 | Invertebrates biomass |

**Table 2**

Fish community in the four rivers. The cyprinid ﬁsh community varied across rivers and was summarised in a single variable.

Ebron Vallanca Palancia Villahermosa

previous studies ﬁsh densities were log(*x* + 1) transformed ([Brosse](#_bookmark45) [and](#_bookmark45) [Lek,](#_bookmark45) [2000;](#_bookmark45) [Fukuda](#_bookmark45) [et al.,](#_bookmark45) [2011b).](#_bookmark45)

Benthic invertebrates were collected with a Hess sampler (0.5 m2) following the International Standard ISO 8265:1988, ofﬁ-

*Salmo trutta Oncorhynchus mykiss Babus haasi Luciobarbus guiraonis Anguilla anguilla*

*Salmo trutta Oncorhynchus mykiss*

*Babus haasi Luciobarbus guiraonis* *Achondrostoma arcasii*

*Salmo trutta Oncorhynchus mykiss*

*Babus haasi Luciobarbus guiraonis Anguilla anguilla*

*Salmo trutta Oncorhynchus mykiss*

*Babus haasi Luciobarbus guiraonis Achondrostoma arcasii*

*Squalius valentinus Anguilla anguilla*

cial version of the European Standard EN 29265 (January 1994). Samples were later identiﬁed to the lowest possible taxonomic level (predominantly at family level), sorted and counted to obtain

the density of invertebrates. Specimens were dried in an oven at 65 ◦C for 24 h and the dry residue was weighed to obtain inver- tebrate biomass. Finally, two additional predictors were derived:

invertebrate richness (i.e. the sum of present taxa in each sam- ple) and invertebrate diversity by applying the Shannon-Weaver diversity index based on the number of individuals per taxa at each sampled HMU ([Table 1).](#_bookmark4)

The river ﬂow was gauged in at least one cross-section using an electromagnetic current metre (Valeport®), and ﬂow velocity was calculated by dividing the ﬂow by the average cross-section area. Elevation and slope were extracted from cartography in a geographic information system, whereas habitat variability was estimated with the Shannon-Weaver diversity index, taking into account the number of habitat types (i.e. number of pools, glides, rifﬂes or rapids) from the visual stratiﬁcation of each study site ([Table 1).](#_bookmark4)

* + - 1. *Biological survey*

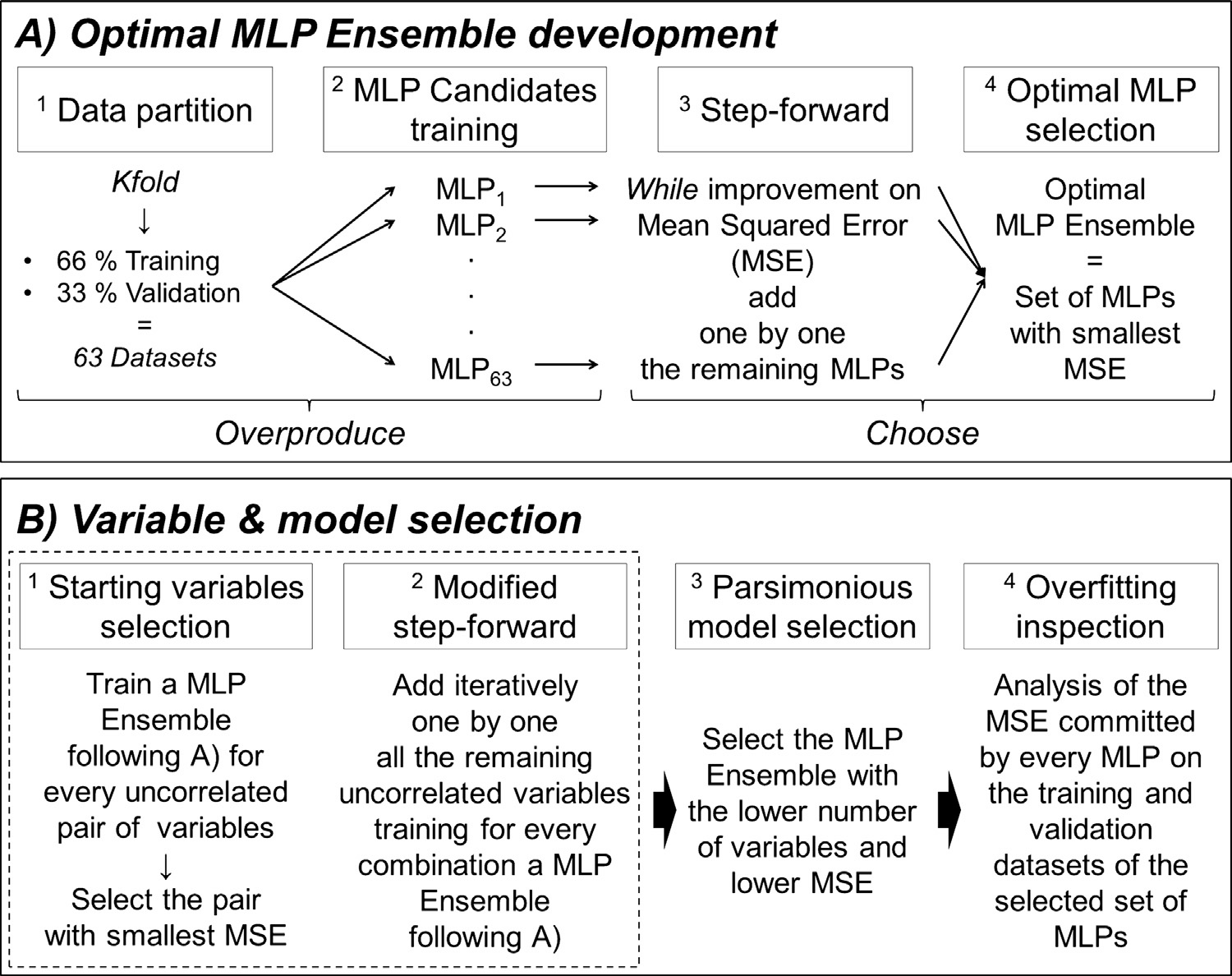
The biological survey was undertaken by electroﬁshing, and all captured ﬁsh species were recorded. In each study site, one slow (i.e. pool or glide) and one fast (i.e. rifﬂe or rapid) HMU were selected and surveyed (3-passes removal) after netting off the HMU. Due to a severe drought, some study sites were dry, resulting in 93 HMUs being sampled. Redﬁn barbel males are mature at approx- imately 45 mm, while females are mature at 100 mm ([Aparicio,](#_bookmark28) [2002),](#_bookmark28) and therefore it was regarded as conservative to consider all specimens larger than 45 mm, resulting in a prevalence of 0.42. No size restrictions were imposed on the remaining ﬁsh species, and thus all the individuals were considered in the data analy- sis. Since the ﬁsh community varied across streams, the cyprinid species were grouped in a single variable ([Table 2),](#_bookmark5) and following

* + 1. *Models’ development*

The physical habitat and biotic models were developed by means of MLP Ensembles ([Hansen](#_bookmark32) [and](#_bookmark32) [Salamon,](#_bookmark32) [1990).](#_bookmark32)

The development of the optimal MLP Ensembles followed the overproduce-and-choose approach. This approach consists of the generation of an initial large pool of MLP candidate classiﬁers (over- produce) whereas the second phase is devoted to select the best performing subset of MLPs (choose). The choose phase was per- formed by means of the step-forward algorithm. Thus, starting from every MLP candidate classiﬁer the best complementary MLP can- didate is iteratively searched until no improvement in the mean squared error (MSE) was achieved ([Fig. 2](#_bookmark6)A).

To render parsimonious models, the optimal input variables’ subsets for both models were also selected by means of the step- forward algorithm. First, the best pair of input variables was determined by developing a MLP Ensemble for every uncorrelated pair following the aforementioned procedure and then this pair became the base for the following step forward variable selection. The algorithm continued until no more variables were available and the selected model was the one with the lowest number of variables and error. Finally, in order to rule out overﬁtting, we visually esti- mated differences between the distributions of the MSE based on the training and validation datasets of the selected MLPs ([Fig. 2](#_bookmark6)B).



**Fig. 2.** Flowchart of the steps followed in the development of the physical habitat and biotic models.

* + - 1. *Multilayer Perceptron Ensemble development*

Building a MLP Ensemble involves training several individual models (MLPs) and combining them to produce aggregated pre- dictions ([Hansen](#_bookmark32) [and](#_bookmark32) [Salamon,](#_bookmark32) [1990).](#_bookmark32) The construction of the individual models (hereafter MLP candidates) was carried out in *R* ([R](#_bookmark59) [Development](#_bookmark59) [Core](#_bookmark59) [Team,](#_bookmark59) [2012)](#_bookmark59) with the package *monmlp* which optimises the model weights using the non-linear minimisation (*nlm)* routine ([Cannon,](#_bookmark16) [2012).](#_bookmark16) The activation functions were the hyperbolic tangent and the linear transformation, while the num- ber of nodes was restricted to improve generalisation ([Özesmi](#_bookmark49) [et al.,](#_bookmark49) [2006)](#_bookmark49) following Eq. (1):

[2006;](#_bookmark61) [Zhou](#_bookmark61) [et al.,](#_bookmark61) [2002),](#_bookmark61) but to our knowledge those sophisticated methods have not been coupled to a variable selection procedure. Consequently, we applied a step-forward selection of the MLP candidates which has been proved to perform similarly to more complex algorithms ([Mun˜ oz-Mas](#_bookmark67) [et al.,](#_bookmark67) [2014).](#_bookmark67) Our step-forward selection was run starting from each of the 63 MLP candidates, searching for the best combination and stopping when no improve- ment was achieved. This was in contrast to the usual step-forward routine where the procedure would start from the best single model.

Since the optimal MLP Ensemble may not include all the MLP

candidates, the observed performance could be affected by over-

number of variables + 1

*N*nodes = max (f1*,* I

2

l'\ (1)

ﬁtting because the selected models may be trained only with some of the training database. Therefore the role of test data was

The aggregated forecast was determined by averaging the indi- vidual predictions of each selected model. Since model training depends on initial conditions, every MLP candidate was optimised ﬁve times with 500 iterations each.

Heterogeneity, or diversity, between MLP candidates is crucial because MLP Ensembles achieve better generalisation when mod- els are complementary ([Opitz,](#_bookmark47) [1999).](#_bookmark47) Several approaches allow for the construction of an MLP Ensemble with heterogeneous candi- dates, such as using different training datasets, architectures or learning methods ([Brown](#_bookmark17) [et al.,](#_bookmark17) [2005).](#_bookmark17) To increase heterogeneity among models, the database was divided in 63 different training and validation datasets corresponding to all possible combinations of 66% of the cases for training, and 33% for validation (i.e. following the *k*-fold approach). Consequently, 63 different MLP candidates were trained for every tested combination of input variables.

Originally the MLP Ensembles included all the developed mod- els ([Hansen](#_bookmark32) [and](#_bookmark32) [Salamon,](#_bookmark32) [1990),](#_bookmark32) but it was promptly demonstrated that active selection of the MLP candidates improved the ﬁnal pre- dictions ([Opitz](#_bookmark48) [and](#_bookmark48) [Shavlik,](#_bookmark48) [1996;](#_bookmark48) [Zhou](#_bookmark48) [et al.,](#_bookmark48) [2002).](#_bookmark48) There are several methods to apply the overproduce-and-choose approach ([Soares](#_bookmark61) [et al.,](#_bookmark61) [2013;](#_bookmark61) [Wang](#_bookmark61) [and](#_bookmark61) [Alhamdoosh,](#_bookmark61) [2013;](#_bookmark61) [Yao](#_bookmark61) [and](#_bookmark61) [Xu,](#_bookmark61)

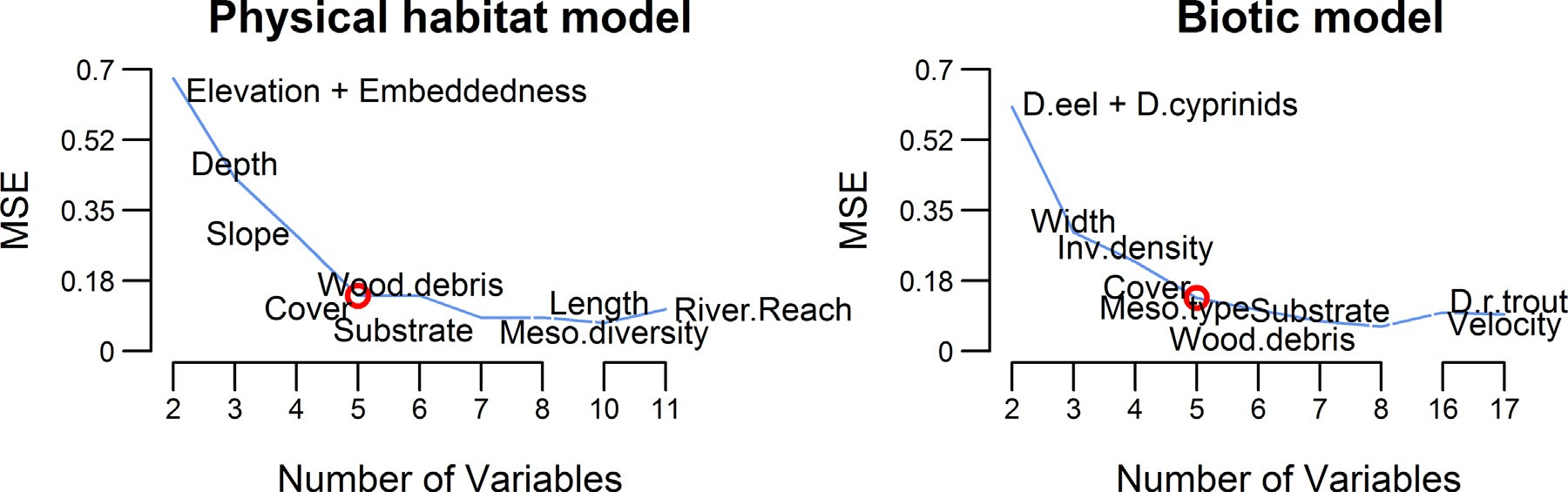
parts

twofold; ﬁrst we applied an a priori *regularisation* method with the early stop regularisation (*sensu* [Ludwig](#_bookmark53) [et al.,](#_bookmark53) [2014)](#_bookmark53) by calculat- ing the MSE on the validation dataset every 100 iterations of the *nlm* routine and then we visually estimated for each selected MLP differences between the distributions of the MSE based on the train- ing and validation datasets of all the selected MLPs. In the case of dissimilar distributions, the number of nodes and the number of iterations ran between calculations of the MSE on the validation dataset were readjusted.

To allow for the comparison with previous studies that either included all trained networks ([Palialexis](#_bookmark51) [et al.,](#_bookmark51) [2011)](#_bookmark51) or based the model selection on a ranking of the individual performances, including the top MLPs ([Watts](#_bookmark68) [and](#_bookmark68) [Worner,](#_bookmark68) [2008),](#_bookmark68) the MSE of the best MLP candidate and of the MLP Ensemble Complete (i.e. the one without any models’ selection) and then for the best ﬁve, ten and ﬁfteen models were calculated and compared with our optimal MLP Ensembles.

* + - 1. *Variable selection*

To identify the most important variables shaping the ecological niche, for both models an input variable subset was selected based



**Fig. 3.** Sequence of the variable selection during the step-forward procedure to develop the MLP Ensemble (from left to right). The plots show the mean squared error, MSE, in function of the number of variables. The circle indicates the selection of the optimal model.

on the step-forward procedure because it has proven computation- ally efﬁcient and tends to result in relatively small input variables’ subsets ([May](#_bookmark60) [et al.,](#_bookmark60) [2011).](#_bookmark60) In contrast to some other approaches (e.g. generalised additive mixed models, [Lin](#_bookmark50) [and](#_bookmark50) [Zhang,](#_bookmark50) [1999)](#_bookmark50) the MLP Ensemble approach does not speciﬁcally allow for the con- sideration of spatial or temporal autocorrelation among training data. To rule out any inﬂuence of study site and sampling year, they were included as input variables ([Table 1).](#_bookmark4) Their absence on the ultimate models would indicate their irrelevance, thus corrob- orating the properness of the data packing. In addition, to render parsimonious models, instead of the usual step-forward procedure that discontinues when no improvement is achieved, the proce- dure was sustained until no more variables were available. The performance of the best model (MLP Ensemble) and the number of variables considered at every iteration were rescaled between 0 and 1 (1 being optimal), with the optimal MLP Ensemble being the one that maximised the sum of both criteria. The step-forward pro- cedure may fail to consider variable interactions and may depend on the variable that was selected ﬁrst. To overcome this limita- tion, one model was developed for each pairwise combination of variables. The best pair of variables was selected as the starting set of variables in the step-forward procedure. Additionally, dur- ing the entire process, neither correlated (*r*2 > 0.5) nor collinear (variable inﬂation factor; *vif* > 5) combinations of variables were considered. Since the input database was a combination of ordi- nal and continuous variables, the function *hetcor* in the package *polycor* ([Fox,](#_bookmark22) [2010)](#_bookmark22) was used to calculate the variables’ correlation (Appendix A1).

* + 1. *Partial dependence plots and uncertainty analysis*

Model reliability and transparency is of major concern for eco- logical modelling ([Austin,](#_bookmark34) [2007;](#_bookmark34) [Guisan](#_bookmark34) [and](#_bookmark34) [Thuiller,](#_bookmark34) [2005;](#_bookmark34) [Özesmi](#_bookmark34) [et al.,](#_bookmark34) [2006)](#_bookmark34) and is fundamental when models are used with exploratory purposes. Therefore, to graphically characterise the relationship between the input variables and the predicted den- sities obtained by the optimal MLP Ensembles, partial dependence plots ([Friedman,](#_bookmark23) [2001)](#_bookmark23) implemented in the package *randomForests* ([Liaw](#_bookmark48) [and](#_bookmark48) [Wiener,](#_bookmark48) [2002)](#_bookmark48) were developed.

The importance of dealing with uncertainty has been stressed as a key challenge in ecological modelling ([Larocque](#_bookmark43) [et al.,](#_bookmark43) [2011).](#_bookmark43) Consequently, partial dependence plots were developed also for every model in the optimal MLP Ensemble, and the function *densregion.normal* in the package *denstrip* ([Jackson,](#_bookmark36) [2008)](#_bookmark36) was used to visually inspect the uncertainty associated to the MLP aggregation in comparison with the input variable distribution.

**Table 3**

Mean squared error (MSE) of the best MLP candidate, the optimal MLP Ensemble, the MLP Ensemble Complete (considering all MLPs; 63) and of the top ﬁve, top ten and top ﬁfteen MLP. The amount of considered networks appears in brackets.

|  |  |  |
| --- | --- | --- |
| Model | Physical habitat model | Biotic model |
| Best MLP candidate | 0.31 (1) | 0.27 (1) |
| Optimal MLP Ensemble | 0.14 (15) | 0.13 (8) |
| MLP Ensemble Complete | 0.95 (63) | 0.31 (63) |
| Top MLP – 1–5 | 0.23 (5) | 0.19 (5) |
| Top MLP – 1–10 | 0.21 (10) | 0.18 (10) |
| Top MLP – 1–15 | 0.18 (15) | 0.18 (15) |

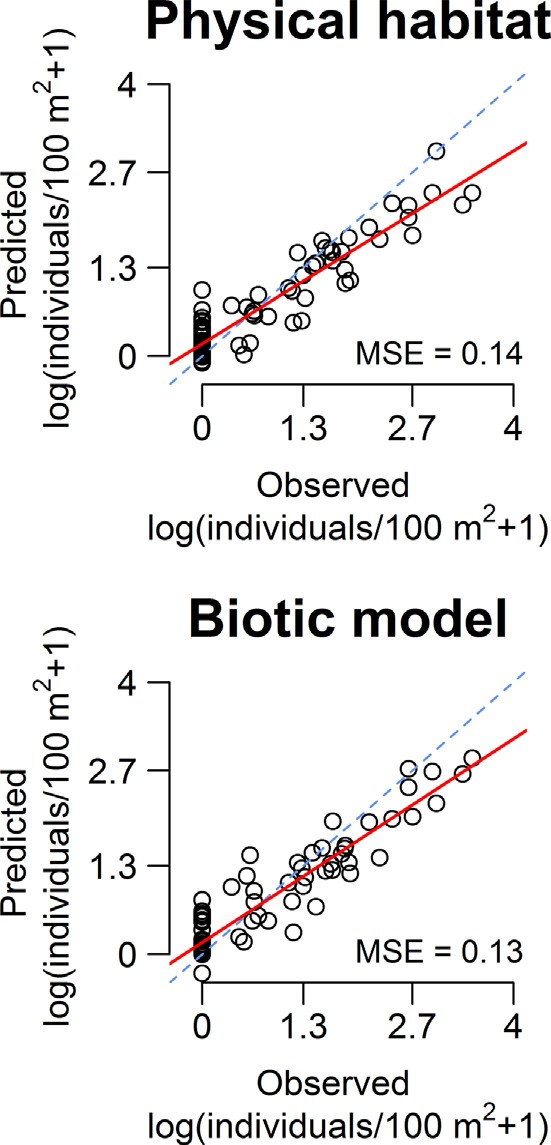
* 1. **Results**
     1. *Training results*

The optimal physical habitat model included ﬁve variables (three nodes): elevation, embeddedness, depth, slope and cover ([Fig. 3)](#_bookmark7) with a maximum correlation of 0.33 and a variable inﬂa- tion factor of 1.41. The optimal biotic model also included ﬁve variables (three nodes): density of eel, cyprinids’ density, width, invertebrates’ density and cover ([Fig. 3)](#_bookmark7) with a maximum corre- lation of 0.38 and variable inﬂation factor of 1.51. In addition, the spatiotemporal correlation was considered negligible since study site and sampling year were not selected as inputs in the ultimate models (i.e. the physical habitat and the biotic models).

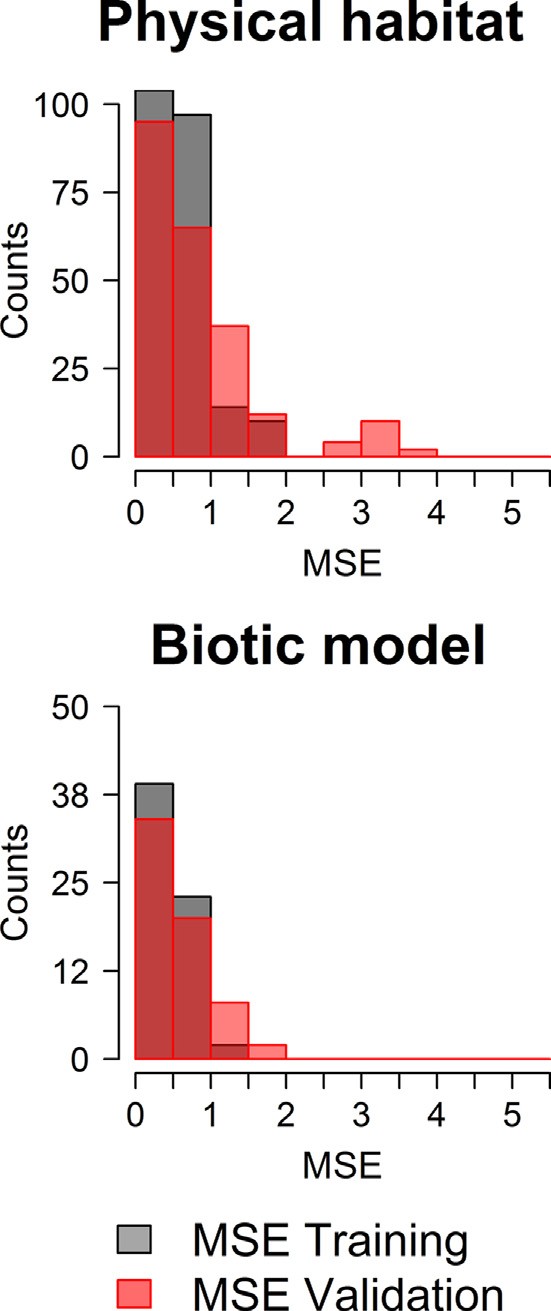
Although the two models showed similar performance (i.e. sim- ilar values of the mean squared error, MSE) and the relatively large amount of zeros in the training dataset slightly biased their out- puts, the biotic model slightly outperformed the physical habitat model with MSE of 0.12 and 0.13 respectively ([Fig. 4](#_bookmark9) and [Table 3).](#_bookmark8) A complete description of the calculated MSE in every iteration is given in Appendix A2.

The physical habitat model selected 15 MLPs. Consequently, 15 training datasets and 15 validation datasets were involved in its development. Cross-evaluation (i.e. the evaluation of every training and validation dataset with every selected MLP candidate) showed that MSEs were distributed equally for the training and the vali- dation datasets ([Fig. 5),](#_bookmark10) and therefore we considered the physical habitat model not overﬁtted. The biotic model selected eight MLPs, with eight training and validation datasets involved in the develop- ment of the selected MLPs. Likewise, the distribution of the training and validation MSE clearly overlapped, and therefore the biotic model was also considered not overﬁtted. The training datasets of the MLP candidates selected within the optimal MLP Ensembles are described in Appendix A3.

The best MLP candidate, the MLP Ensemble Complete (i.e. considering all sixty three MLPs) and the top ﬁve, top ten and top



**Fig. 4.** Relation between the observed and predicted values of the optimal MLPs. The optimal physical habitat and biotic models show transformed output.



**Fig. 5.** Frequency analysis of the mean squared error (MSE) of the selected MLP candidates based on the corresponding training and validation datasets.

ﬁfteen MLPs yielded higher MSEs than the optimal MLP Ensem- bles. The highest difference appeared between the MLP Ensemble Complete and the optimal physical habitat model ([Table 3).](#_bookmark8)

* + 1. *Partial dependence plots – physical habitat model*

The optimal physical habitat model showed a unimodal response between redﬁn barbel density and elevation, with a max- imum density at 738 m above sea level. Embeddedness showed an almost ﬂat trend but an exponential increase from 75% onwards. Depth showed a steep positive linear trend, thus suggesting the major impact among the selected variables, whereas slope and cover were negatively and almost linearly related to redﬁn bar- bel density. As expected, uncertainty was higher at the extremes of the variables’ distributions and therefore trends at these extreme values could be unreliable ([Fig. 6).](#_bookmark11)

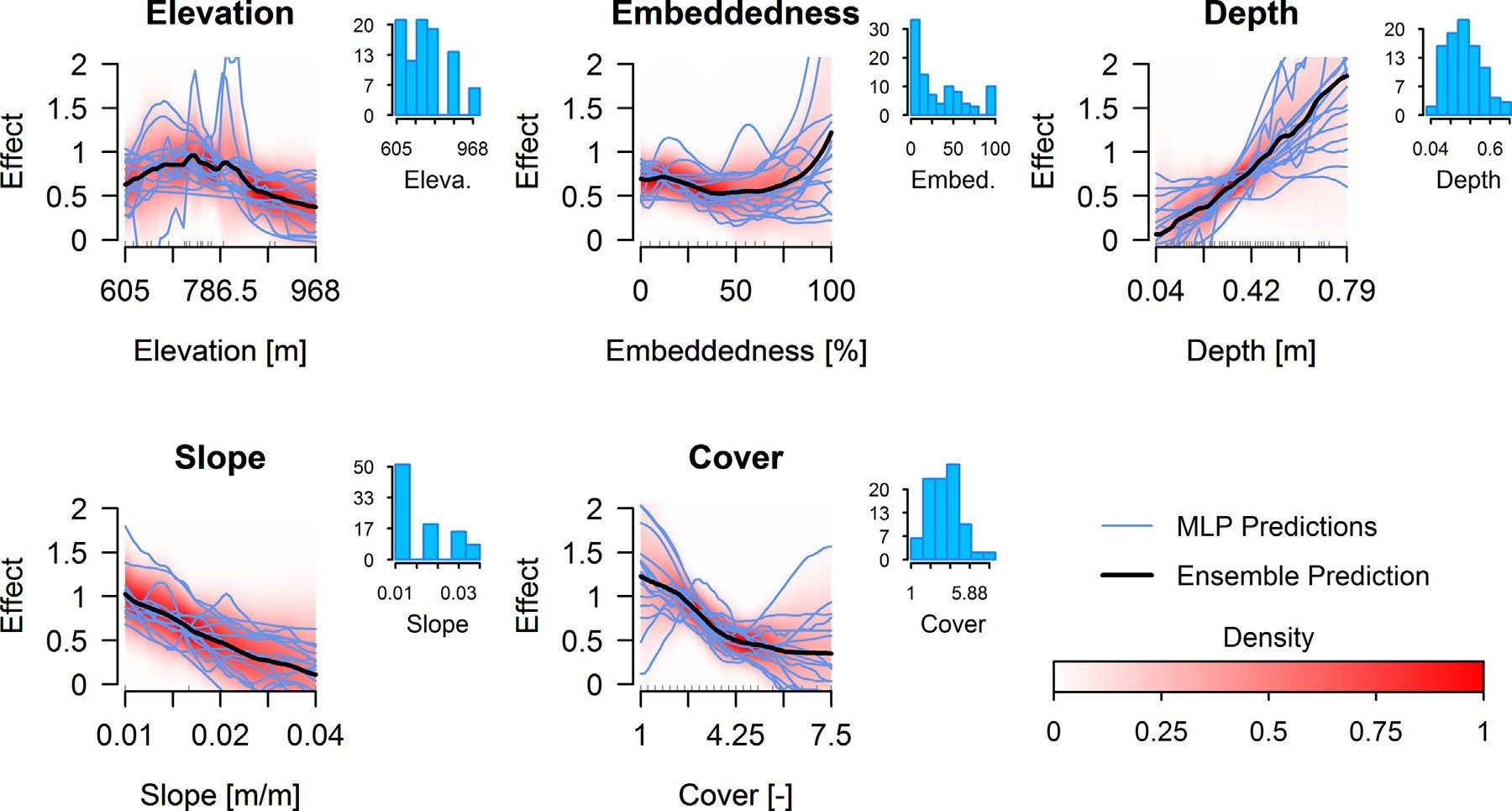
* + 1. *Partial dependence plots – biotic model*

The optimal biotic model showed a positive linear rela- tion between eel and redﬁn barbel densities. Cyprinids’ den- sity presented a unimodal response with the peak around 4 individuals/100 m2. Width showed an almost positive linear inﬂuence on redﬁn barbel density whereas invertebrates’ density presented a unimodal response inﬂecting at 21718 individuals/m2. Likewise the physical habitat model cover presented a linear trend but with smaller uncertainty and slope. Uncertainty was higher than in the physical habitat model although it presented a simi- lar pattern with the extreme values being more uncertain than the central part of the input variables’ distributions ([Fig. 7).](#_bookmark12)

* 1. **Discussion**
     1. *MLP Ensemble development*

Our results indicated that the MLP Ensemble paradigm can be considered proﬁcient to model the ecological niche of freshwa- ter ﬁsh species, in line with previous studies that modelled ﬁsh density with neural networks ([Baran](#_bookmark35) [et al.,](#_bookmark35) [1996;](#_bookmark35) [Brosse](#_bookmark35) [and](#_bookmark35) [Lek,](#_bookmark35) [2000;](#_bookmark35) [Laffaille](#_bookmark35) [et al.,](#_bookmark35) [2003).](#_bookmark35) The presented optimal models also out- performed any single MLP, which agrees with previous research ([Palialexis](#_bookmark51) [et al.,](#_bookmark51) [2011).](#_bookmark51) We also demonstrated that active selection of MLP candidates to create an optimal subset can further improve MLP Ensembles’ performance. This is consistent with [Zhou’s](#_bookmark71) [et al.](#_bookmark71) [(2002)](#_bookmark71) statement that “*many could be better than all*”. Moreover, our candidates’ selection approach resulted in a better performance in contrast with the selection approach based on the individual performance (i.e. top ﬁve, ten and ﬁfteen). We recommend this procedure in contrast to previous studies that selected the best subset based on the individual performance of the MLPs ([Watts](#_bookmark68) [and](#_bookmark68) [Worner,](#_bookmark68) [2008).](#_bookmark68) However, our step-forward process is determined by the ﬁrst selected model, and despite the fact that the proce- dure started from every single neural network, the possibility to get stuck at a local minimum exists. Therefore, untested combinations of models could outperform those obtained by the step-forward algorithm.

Genetic algorithms may overcome the aforementioned con- straints ([Soares](#_bookmark61) [et al.,](#_bookmark61) [2013;](#_bookmark61) [Wang](#_bookmark61) [and](#_bookmark61) [Alhamdoosh,](#_bookmark61) [2013).](#_bookmark61) There are successful applications of genetic algorithms in variable selec- tion procedures ([May](#_bookmark60) [et al.,](#_bookmark60) [2011;](#_bookmark60) [Olden](#_bookmark60) [et al.,](#_bookmark60) [2008)](#_bookmark60) and also within the selection of optimal MLP Ensembles ([Soares](#_bookmark61) [et al.,](#_bookmark61) [2013;](#_bookmark61) [Wang](#_bookmark61) [and](#_bookmark61) [Alhamdoosh,](#_bookmark61) [2013).](#_bookmark61) Although the use of genetic algorithms for both variable selection and MLP candidates’ selection could expo- nentially increase the computation effort, this approach is certainly promising and should be the subject of future research.



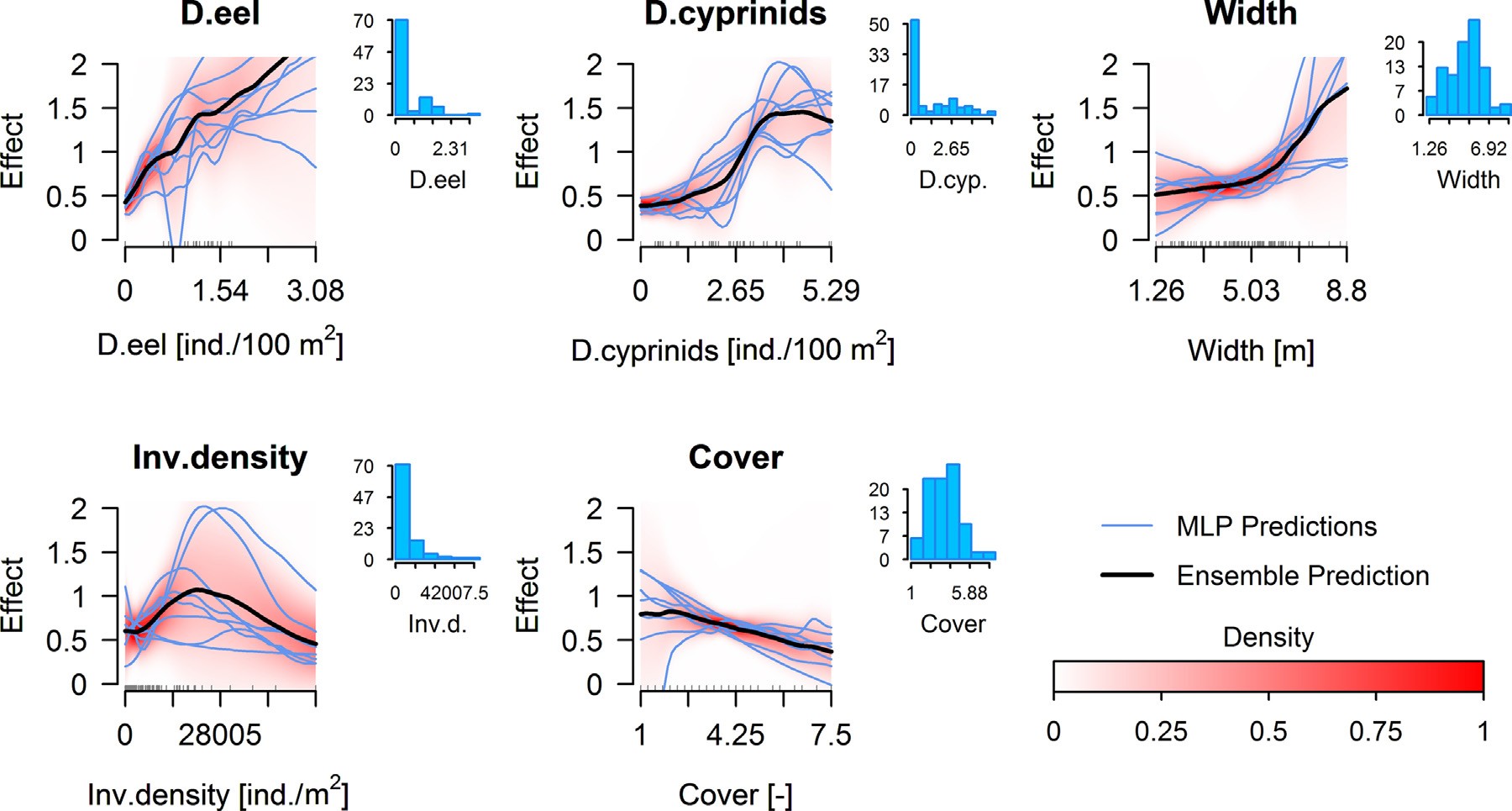
**Fig. 6.** Partial dependence plots of the physical habitat model (black line). Light lines (blue lines on the e-version) correspond to the partial dependence plot of every selected MLP candidate. Faded background corresponds to uncertainty analysis based on mean and standard deviation of predictions, and the darker the colour the smaller the uncertainty.

The effect of the relatively large amount of zeros in the train- ing dataset was a remarkable issue and slightly biasing models’ outputs. There are speciﬁc techniques in count data modelling to deal with excess of zeros with either parametric ([Mullahy,](#_bookmark37) [1986;](#_bookmark37) [Lambert,](#_bookmark37) [1992)](#_bookmark37) and non-parametric ([Liu](#_bookmark52) [and](#_bookmark52) [Chan,](#_bookmark52) [2010)](#_bookmark52) responses. Certainly the comparison of the capability of MLP Ensembles and these techniques in ecological modelling would be of interest, although these techniques by deﬁnition do not eas- ily account for variable interactions and thus do not easily assure better performance. Nevertheless, our results were considered sat- isfactory since they provided an acceptable balance between model complexity, performance and computational effort, and they were

devoted to explore rather than to predict ﬁsh density in further analysis.

* + 1. *Ecological relevance of the physical habitat model*

The optimal physical habitat model included ﬁve variables: elevation, slope, depth, embeddedness and cover. Elevation is broadly accepted as a proximal predictor of temperature ([Elith](#_bookmark19) [and](#_bookmark19) [Leathwick,](#_bookmark19) [2009),](#_bookmark19) and consequently we considered that it may reﬂect the effect of climate on redﬁn barbel distribution. Sim- ilarly, the negative trend of redﬁn barbell density versus slope agrees with its preference for middle-to-upper stream reaches of



**Fig. 7.** Partial dependence plots of the biotic model (black line). Light lines (blue lines on the e-version) correspond to the partial dependence plot of every selected MLP candidate. Faded background corresponds to uncertainty analysis based on mean and standard deviation of predictions, and the darker the colour the smaller the uncertainty.

mountainous rivers ([Perea](#_bookmark56) [et al.,](#_bookmark56) [2011).](#_bookmark56) The positive impact of depth corresponds with previous studies that considered the redﬁn barbel a pool dweller ([Aparicio](#_bookmark30) [and](#_bookmark30) [De](#_bookmark30) [Sostoa,](#_bookmark30) [1999).](#_bookmark30) Our results may also suggest the importance of backwaters or stagnated areas as resting habitats. This could also explain the positive relationship between redﬁn barbell density and embeddedness, since pool substrates are generally more embedded. Despite the negative relationship between cover and redﬁn barbel density in our study, some authors classiﬁed the redﬁn barbel as a cover-oriented ﬁsh ([Grossman](#_bookmark29) [and](#_bookmark29) [De](#_bookmark29) [Sostoa,](#_bookmark29) [1994).](#_bookmark29) [Aparicio](#_bookmark28) [(2002)](#_bookmark28) reported the active use of cover in an ephemeral river but related its use to the absence of deep pools in this speciﬁc river, rather than to a redﬁn barbel preference for cover. Our study suggests that in more complex river systems with well-developed pool-rifﬂe patterns, redﬁn barbel may tend to avoid excessive cover complexity.

* + 1. *Ecological relevance of the biotic model*

In contrast to previous studies ([Vezza](#_bookmark66) [et al.,](#_bookmark66) [2015;](#_bookmark66) [Watts](#_bookmark66) [and](#_bookmark66) [Worner,](#_bookmark66) [2008)](#_bookmark66) where the combination of physical habitat and biotic variables outperformed the model developed only with physical habitat variables, our biotic model did not signiﬁcantly perform better than the physical habitat model. Moreover, uncertainty was higher in the biotic model, which underlines the previously reported complexity related to the assessment of biotic interac- tions ([Leathwick](#_bookmark46) [and](#_bookmark46) [Austin,](#_bookmark46) [2001).](#_bookmark46) Although larger datasets may reduce this uncertainty, our study nevertheless suggests interesting and plausible relationships. In addition, the observed associations between biotic variables and red ﬁn density agreed with the eco- logical gradient theory because responses were quasi-linear or unimodal ([Austin,](#_bookmark34) [2007).](#_bookmark34)

The biotic model demonstrated a positive association between redﬁn barbel and European eel, which conﬁrms the work of [Laffaille](#_bookmark40) [et al.](#_bookmark40) [(2003).](#_bookmark40) They modelled eel habitat suitability in a small coastal catchment with a single MLP. Eels were more abundant in deep and low ﬂow shaded areas without aquatic vegetation ([Laffaille](#_bookmark40) [et al.,](#_bookmark40) [2003).](#_bookmark40) Such a pattern broadly concurs with the requirements of redﬁn barbel. Also the suggested relationship between redﬁn bar- bel and cyprinids corresponds with previous studies on the Iberian Peninsula that reported the presence of multi-species shoals as well as an overlap in microhabitat use ([Martínez-Capel](#_bookmark55) [et al.,](#_bookmark55) [2009).](#_bookmark55) Fish schooling beneﬁts include the enhancement of hydrodynamics and the protection against predators ([Landa,](#_bookmark44) [1998).](#_bookmark44) Moreover, simi- lar positive interactions with cyprinid species have been reported for Iberian chub (*Squalius pyrenaicus*, Günther, 1868) and eastern Iberian barbel (*Luciobarbus guiraonis*, Steindachner, 1866) ([Vezza](#_bookmark66) [et al.,](#_bookmark66) [2015).](#_bookmark66) This indicates that restoration actions focused on redﬁn barbel could also result in habitat enhancement for other cyprinid species.

Although previous work positively correlated invertebrate den- sity to ﬁsh density ([Mas-Martí](#_bookmark57) [et al.,](#_bookmark57) [2010),](#_bookmark57) our results show a maximal redﬁn barbel density at 21718 individuals/m2. This could be related to food availability. However, the preferred prey invertebrates of redﬁn barbel (i.e. *Chironomidae*, *Ephemeroptera* and *Trichoptera* following [Miranda](#_bookmark62) [et al.,](#_bookmark62) [2005)](#_bookmark62) were strongly corre- lated with the invertebrate density applied in our model. Therefore, we attributed the avoidance of the higher invertebrate densities to a habitat correlation. Previous studies suggested that in Mediter- ranean rivers, with very unstable climatic conditions, rifﬂes tend to host higher invertebrate density than pools ([Bonada](#_bookmark41) [et al.,](#_bookmark41) [2006).](#_bookmark41) Therefore the decrement of the partial dependence plot could be showing the necessity for larger depth rather than a preference for intermediate invertebrate densities. Nevertheless, the discrepancy

between our results and the literature can also be a consequence of the applied model complexity (e.g. linear vs. non-linear models), and the impact of model complexity should be thoroughly analysed

in further studies modelling the relationship between invertebrate and ﬁsh density. The biotic model also selected two physical habitat variables, cover index and width, and the relationship with cover index being similar to that shown in the physical habitat model. The positive association between redﬁn barbel density and width may be attributed to the negative correlation between width and elevation, in line with the aforementioned redﬁn barbel preference for middle-to-upper stream reaches, but with a slightly different response because the study encompassed four different rivers.

Although the biotic model appears to suggest interactions between redﬁn barbel and other species, signiﬁcant positive or neg- ative correlations between species does not imply a causative effect ([Wisz](#_bookmark69) [et al.,](#_bookmark69) [2013).](#_bookmark69) A simple correlation does not mandatorily corre- spond to any species interaction, neither mutualism nor facilitation, and therefore further research should clarify the true impact of species interactions. Furthermore, changes in the habitat available may result in a substantial increment in the competition between species ([Wisz](#_bookmark69) [et al.,](#_bookmark69) [2013).](#_bookmark69) The cyprinids density partial depend- ence plot showed a decrement at the tail of the curve. Therefore, in spite of being uncertain, it could suggest that, under different habitat conditions than those in our study, the positive interaction with cyprinid ﬁsh species may become habitat competition, thus emphasising the necessity of close monitoring in the near future to avoid ecological loss.

* + 1. *Model uncertainty*

Relatively few studies address uncertainty in ecological mod- elling and its effects on model predictions and decision making ([Elith](#_bookmark19) [and](#_bookmark19) [Leathwick,](#_bookmark19) [2009).](#_bookmark19) In accordance with previous studies ([Peters](#_bookmark58) [et al.,](#_bookmark58) [2009),](#_bookmark58) the largest uncertainty tended to appear in the regions of the input variables that were poorly represented in the training database. In contrast to the high uncertainty demon- strated by the different MLP candidate predictions, the optimal MLP Ensembles produced sound and smoother partial dependence plots that allowed general trends to be derived from a wide range of model outputs. This has been stimulated by three approaches applied in our study. First, the bias and variance dilemma ([Geman](#_bookmark25) [et al.,](#_bookmark25) [1992)](#_bookmark25) was addressed by limiting model complexity (i.e. limiting the number of nodes and variables), leading to less com- plex models than in previous studies ([Dedecker](#_bookmark18) [et al.,](#_bookmark18) [2004;](#_bookmark18) [Lek](#_bookmark18) [et al.,](#_bookmark18) [1996).](#_bookmark18) Second, the early stop regularisation considered the errors committed on the training and validation datasets ([Ludwig](#_bookmark53) [et al.,](#_bookmark53) [2014).](#_bookmark53) Third and most importantly, we assessed overﬁtting by checking whether species responses to environmental variables were consistent with the ecological gradient theory ([Austin,](#_bookmark34) [2007).](#_bookmark34) Inconsistent model results would have suggested a more restric- tive modelling approach by limiting model complexity or adjusting the early stop parameters. Uncertainty could also arise when sam- ples from different periods are combined, since ﬁsh density is a density-dependent phenomenon ([Mas-Martí](#_bookmark57) [et al.,](#_bookmark57) [2010);](#_bookmark57) how- ever, the sampling year was not selected as an important variable in the optimal MLP Ensembles. Consequently, the results from our study suggest that temporal packing can be considered admissible when focusing on a short time span.

* 1. **Conclusions**

The MLP Ensembles have demonstrated satisfactory perfor- mance and generated predictions in line with the ecological gradient theory. As such, the models provide a better insight into the ecological niche of redﬁn barbel by complementing pre- vious studies. We not only demonstrated that MLP Ensembles may outperform single MLPs, but also that the active selection of MLP candidates to create an optimal subset of models can

further improve model performance. The physical habitat model conﬁrmed the redﬁn barbel preference for middle-to-upper river segments, but not in higher and steeper reaches. The importance of depth conﬁrms that redﬁn barbel prefer pool-type habitats, which emphasises the vulnerability of the species to reduced ﬂows. Although the biotic model showed higher uncertainty, it suggested that redﬁn barbel, European eel and the considered cyprinid species have similar habitat requirements. Due to its high predictive per- formance and its ability to deal with model uncertainty, the MLP Ensemble paradigm is a promising tool for ecological modelling or habitat suitability prediction to assess environmental ﬂows.

**Acknowledgements**

This study was funded by the Spanish Ministry of Economy and Competitiveness with the project SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and the Universitat Politècnica de València, through the project UPPTE/2012/294 (PAID-06-12). Additionally, the authors would like to thank the help of the Conselleria de Territori i Vivenda (Generalitat Valenciana) and the Confederación Hidrográﬁca del Júcar (Spanish government) which provided envi- ronmental data. The authors are indebted to all the colleagues who collaborated in the ﬁeld data collection and the text adequacy; without their help this paper would have not been possible. Last but not least, the authors would like to speciﬁcally thank E. Aparicio and A.J. Cannon, the former because he selﬂessly provided the bibli- ography about the redﬁn barbel and the latter because he patiently explained the ‘ins and outs’ of the *monmlp* package.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at [http://dx.doi.org/10.1016/j.ecolmodel.2015.](http://dx.doi.org/10.1016/j.ecolmodel.2015.04.025) [04.025](http://dx.doi.org/10.1016/j.ecolmodel.2015.04.025)

**References**

Alcaraz-Hernández, J.D., Martínez-Capel, F., Peredo, M., Hernández-Mascarell, A., 2011. [Mesohabitat](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [heterogeneity](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [four](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [mediterranean](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [streams](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [the](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [Jucar](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [river](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [basin](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [(Eastern](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [Spain).](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [Limnetica](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [30](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [(2),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005) [15–363.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0005)

Aparicio, E., 2002. [Ecologia](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [del](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [barb](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [cua-roig](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [(*Barbus*](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010)[*haasi*)](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [i](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [avaluació](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [del](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [seu](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [estat](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [de](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [conservació](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [a](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Catalunya.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Programa](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [de](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Doctorat](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [de](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Biologia](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Animal](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [I](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [–](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Zoologia](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010)

[–](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [Bienni](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [1991–1993.,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [pp.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [173](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010) [(Catalan).](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0010)

Aparicio, E., De Sostoa, A., 1999. Pattern of movements of adult *Barbus haasi* in a small Mediterranean stream. J. Fish Biol. 55 (5), 1086–1095, <http://dx.doi.org/> 10.1006/jfbi.1999.1109

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22 (1), 42–47, <http://dx.doi.org/10.1016/j.tree.2006.09.010>

Austin, M., 2007. Species distribution models and ecological theory: a critical assess- ment and some possible new approaches. Ecol. Model. 200 (1–2), 1–19, http:// dx.doi.org/10.1016/j.ecolmodel.2006.07.005

Baran, P., Lek, S., Delacoste, M., Belaud, A., 1996. Stochastic models that predict trout population density or biomass on a mesohabitat scale. Hydrobiologia 337 (1–3), 1–9, <http://dx.doi.org/10.1007/BF00028502>

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., et al., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol. Model. 222 (11), 1810–1819, [http://dx.](http://dx/) doi.org/10.1016/j.ecolmodel.2011.02.011

Bianco, P.G., 1998. Diversity of Barbinae ﬁshes in southern Europe with description of a new genus and a new species (Cyprinidae). Ital. J. Zool. 65 (Suppl. 1), 125–136, <http://dx.doi.org/10.1080/11250009809386804>

Bonada, N., Rieradevall, M., Prat, N., Resh, V.H., 2006. Benthic macroinvertebrate assemblages and macrohabitat connectivity in Mediterranean-climate streams of northern California. J. N. Am. Benthol. Soc. 25 (1), 32–43, [http://dx.doi.org/10.](http://dx.doi.org/10) 1899/0887-3593(2006)25[32:bmaamc]2.0.co;2

Boulangeat, I., Gravel, D., Thuiller, W., 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abun- dances. Ecol. Lett. 15 (6), 584–593, [http://dx.doi.org/10.1111/j.1461-0248.2012.](http://dx.doi.org/10.1111/j.1461-0248.2012) 01772.x

Brosse, S., Lek, S., 2000. Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques. Freshwater Biol. 44 (3), 441–452, <http://dx.doi.org/> 10.1046/j.1365-2427.2000.00580.x

Brown, G., Wyatt, J., Harris, R., Yao, X., 2005. Diversity creation methods: a survey and categorisation. Inf. Fusion 6 (1), 5–20, [http://dx.doi.org/10.1016/j.inffus.2004.](http://dx.doi.org/10.1016/j.inffus.2004)

04.004

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Change BioI. 16 (4), 1145–1157, <http://dx.doi.org/10.1111/j.1365-2486.2009.02000.x>

Cannon, A.J., 2012. [monmlp:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Monotone](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Multi-layer](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Perceptron](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Neural](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Network.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [R](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Package](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [Version](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070) [1.1.2.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0070)

Crisci, C., Ghattas, B., Perera, G., 2012. A review of supervised machine learning algo- rithms and their applications to ecological data. Ecol. Model. 240 (0), 113–122, <http://dx.doi.org/10.1016/j.ecolmodel.2012.03.001>

Dedecker, A.P., Goethals, P.L.M., Gabriels, W., De Pauw, N., 2004. Optimization of Arti- ﬁcial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Ecol. Model. 174 (1–2), 161–173, <http://dx.doi.org/10.1016/j.ecolmodel.2004.01.003>

Dolloff, C.A., Hankin, D.G., Reeves, G.H., 1993. [Basinwide](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Estimation](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Habitat](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Fish](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Populations](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Streams](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Gen.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Tech.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Rep.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [SE-83.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Southeastern](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Forest](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Experi-](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [ment](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Station,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [Asheville,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [NC,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085) [USA.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0085)

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Ann. Rev. Ecol. Evol. Syst. 40, 677–697, <http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159>

Ferreira, T., Oliveira, J., Caiola, N., de Sostoa, A., Casals, F., Cortes, R., et al., 2007. Eco- logical traits of ﬁsh assemblages from Mediterranean Europe and their responses to human disturbance. Fish. Manage. Ecol. 14 (6), 473–481, [http://dx.doi.org/10.](http://dx.doi.org/10) 1111/j.1365-2400.2007.00584.x

Figuerola, B., Maceda-Veiga, A., de Sostoa, A., 2012. Assessing the effects of sewage efﬂuents in a Mediterranean creek: ﬁsh population features and biotic indices. Hydrobiologia 694 (1), 75–86, <http://dx.doi.org/10.1007/s10750-012-1132-y>

Fox, J., 2010. [polycor:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [Polychoric](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [Polyserial](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [Correlations.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [R](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [Package](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [Version](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) [0.7-8.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0105) Freyhof, J., Brooks, E., 2011. [European](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Red](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [List](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Freshwater](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Fishes](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Luxembourg.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110)

[Publications](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Ofﬁce](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [the](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [European](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Union,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110) [Luxembourg.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0110)

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine.

Ann. Stat. 29 (5), 1189–1232, <http://dx.doi.org/10.1214/aos/1013203451> Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M., 2013. Habitat

prediction and knowledge extraction for spawning European grayling (*Thymal- lus thymallus* L.) using a broad range of species distribution models. Environ. Model. Softw. 47, 1–6, <http://dx.doi.org/10.1016/j.envsoft.2013.04.005>

Fukuda, S., Mouton, A.M., De Baets, B., 2011a. 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

Fukuda, S., De Baets, B., Mouton, A.M., Waegeman, W., Nakajima, J., Mukai, T., et al., 2011b. Effect of model formulation on the optimization of a genetic Takagi–Sugeno fuzzy system for ﬁsh habitat suitability evaluation. Ecol. Model. 222 (8), 1401–1413, <http://dx.doi.org/10.1016/j.ecolmodel.2011.01.023>

García de Jalón, D., Schmidt, G., 1995. [Manual](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [práctico](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [para](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [la](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [gestión](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [sostenible](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [de](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [la](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [pesca](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [ﬂuvial.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [Madrid,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [Spain](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [(in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135) [Spanish).](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0135)

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4 (1), 1–58, [http://dx.doi.org/10.1162/neco.1992.4.1.](http://dx.doi.org/10.1162/neco.1992.4.1)

1

Gevrey, M., Dimopoulos, I., Lek, S., 2006. Two-way interaction of input variables in the sensitivity analysis of neural network models. Ecol. Model. 195 (1–2), 43–50, <http://dx.doi.org/10.1016/j.ecolmodel.2005.11.008>

Grossman, G.D., De Sostoa, A., 1994. Microhabit use by ﬁsh in the upper Rio Matar- rana, Spain, 1984–1987. Ecol. Freshw. Fish 3 (4), 141–152, [http://dx.doi.org/10.](http://dx.doi.org/10) 1111/j.1600-0633.1994.tb00016.x

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.](http://dx.doi.org/10.1111/j) 1461-0248.2005.00792.x

Hansen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal.

12 (10), 993–1001, <http://dx.doi.org/10.1109/34.58871>

Hopkins II, R.L., Burr, B.M., 2009. Modeling freshwater ﬁsh distributions using mul- tiscale landscape data: a case study of six narrow range endemics. Ecol. Model. 220 (17), 2024–2034, <http://dx.doi.org/10.1016/j.ecolmodel.2009.04.027>

Mullahy, J., 1986. Speciﬁcation and testing of some modiﬁed count data models. J. Econometr. 33 (3), 341–365, <http://dx.doi.org/10.1016/0304-4076(86)90002-3>

Jackson, C.H., 2008. Displaying uncertainty with shading. Am. Stat. 4 (62), 340–347, <http://dx.doi.org/10.1198/000313008X370843>

Kemp, S.J., Zaradic, P., Hansen, F., 2007. An approach for determining relative input parameter importance and signiﬁcance in artiﬁcial neural networks. Ecol. Model. 204 (3–4), 326–334, <http://dx.doi.org/10.1016/j.ecolmodel.2007.01.009>

Laffaille, P., Feunteun, E., Baisez, A., Robinet, T., Acou, A., Legault, A., et al., 2003. Spatial organisation of European eel (*Anguilla anguilla* L.) in a small catchment. Ecol. Freshw. Fish 12 (4), 254–264, [http://dx.doi.org/10.1046/j.1600-0633.2003.](http://dx.doi.org/10.1046/j.1600-0633.2003) 00021.x

Lambert, D., 1992. Zero-inﬂated Poisson regression, with an application to defects in manufacturing. Technometrics. 34, 11–14, <http://dx.doi.org/10.2307/1269547>

Landa, J.T., 1998. Bioeconomics of schooling ﬁshes: selﬁsh ﬁsh, quasi-free riders, and other ﬁshy tales. Environ. Biol. Fish. 53 (4), 353–364, <http://dx.doi.org/10.1023/> A:1007414603324

Larocque, G.R., Mailly, D., Yue, T.-., Anand, M., Peng, C., Kazanci, C., et al., 2011. Com- mon challenges for ecological modelling: synthesis of facilitated discussions held at the symposia organized for the 2009 conference of the International Society for Ecological Modelling in Quebec City, Canada (October 6–9, 2009). Ecol. Model. 222 (14), 2456–2468, [http://dx.doi.org/10.1016/j.ecolmodel.2010.](http://dx.doi.org/10.1016/j.ecolmodel.2010)

12.017

Leathwick, J.R., Austin, M.P., 2001. Competitive interactions between tree species in New Zealand’s old-growth indigenous forests. Ecology 82 (9), 2560–2573, <http://dx.doi.org/10.2307/2679936>

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Appli- cation of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (1), 39–52, <http://dx.doi.org/10.1016/0304-3800(95)00142->

5

Liaw, A., Wiener, M., 2002. [Classiﬁcation](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [regression](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [by](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [randomForest.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [R](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [News](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [3](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [(2),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215) [18–22.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0215)

Lin, X., Zhang, D., 1999. Inference in generalized additive mixed models by using smoothing splines. J. R. Stat. Soc. Ser. B: Stat. Methodol. 61 (2), 381–400, http:// dx.doi.org/10.1111/1467-9868.00183

Liu, H., Chan, K., 2010. [Introducing](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [COZIGAM:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [an](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [R](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [package](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [for](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [unconstrained](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [constrained](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [zero-inﬂated](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [generalized](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [additive](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [model](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [analysis.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [J.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [Stat.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [Software](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [35](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [(11),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225) [1–26.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0225)

Ludwig, O., Nunes, U., Araujo, R., 2014. Eigenvalue decay: a new method for neu- ral network regularization. Neurocomputing 124, 33–42, [http://dx.doi.org/10.](http://dx.doi.org/10) 1016/j.neucom.2013.08.005

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

Mas-Martí, E., García-Berthou, E., Sabater, S., Tomanova, S., Mun˜ oz, I., 2010. Compar- ing ﬁsh assemblages and trophic ecology of permanent and intermittent reaches in a Mediterranean stream. Hydrobiologia 657 (1), 167–180, <http://dx.doi.org/> 10.1007/s10750-010-0292-x

May, R., Dandy, G., Maier, H., 2011. [Review](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Input](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Variable](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Selection](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Meth-](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [ods](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [for](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Artiﬁcial](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Neural](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Networks.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [In:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Suzuki,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [K.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [(Ed.),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Artiﬁcial](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Neural](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Networks](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [–](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Methodological](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Advances](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Biomedical](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [Applications.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [InTech,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [p.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245) [362.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0245)

Miranda, R., Díez-León, M., Escala, M.C., 2005. [Length](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [relationships](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [cyprinid](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [prey](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [diet](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [analysis](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [Eurasian](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [otter](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [*Lutra*](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250)[*lutra*](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250)[in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [Mediterranean](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [habitats.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [Folia](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [Zool.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [54](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [(4),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250) [443–447.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0250)

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, <http://dx.doi.org/> 10.1016/j.envsoft.2010.12.001

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2010. Ecological relevance of perfor- mance criteria for species distribution models. Ecol. Model. 221 (16), 1995–2002, <http://dx.doi.org/10.1016/j.ecolmodel.2010.04.017>

Mun˜ oz-Mas, R., Alcaraz-Hernández, J.D., Martínez-Capel, F., 2014. [Multilayer](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Per-](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [ceptron](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Ensembles](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [(MLP](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Ensembles)](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [in](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [modelling](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [microhabitat](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [suitability](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [for](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [freshwater](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [ﬁsh.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [In:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [XVII](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Congreso](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Espa](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265)n[˜ ol](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [sobre](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Tecnologías](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [y](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Lógica](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Fuzzy](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [(ESTYLF](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [2014),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Zaragoza,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [Spain,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [pp.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265) [609–614.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0265)

Mun˜ oz-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.](http://dx.doi.org/10.1016/j.scitotenv) 2012.07.074

Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: a randomiza- tion 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

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, <http://dx.doi.org/10.1086/> 587826

Opitz, D.W., 1999. [Feature](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [selection](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [for](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [ensembles.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [In:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Proceedings](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [the](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [1999](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [16th](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [National](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Conference](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [on](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Artiﬁcial](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Intelligence](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [(AAAI-99),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [11th](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Innovative](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Applications](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [of](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Artiﬁcial](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Intelligence](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Conference](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [(IAAI-99),](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [Orlando,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [FL,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [USA,](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [pp.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285) [379–384.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0285)

Opitz, D.W., Shavlik, J.W., 1996. Actively searching for an effective Neural Net- work Ensemble. Connect. Sci. 8 (3–4), 337–353, <http://dx.doi.org/10.1080/>

095400996116802

Özesmi, S.L., Tan, C.O., Özesmi, U., 2006. Methodological issues in building, training, and testing artiﬁcial neural networks in ecological applications. Ecol. Model. 195 (1–2), 83–93, <http://dx.doi.org/10.1016/j.ecolmodel.2005.11.012>

Palialexis, A., Georgakarakos, S., Karakassis, I., Lika, K., Valavanis, V.D., 2011. Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics and regression models. Hydrobiologia 670 (1), 165–188, <http://dx.doi.org/10.1007/s10750-011-0676-6>

Park, Y., Chon, T., 2007. Biologically-inspired machine learning implemented to ecological informatics. Ecol. Model. 203 (1–2), 1–7, <http://dx.doi.org/10.1016/> j.ecolmodel.2006.05.039

Perea, S., Garzón, P., González, J.L., Almada, V.C., Pereira, A., Doadrio, I., 2011. New distribution data on Spanish autochthonous species of freshwater ﬁsh. Graellsia 67 (1), 91–102, <http://dx.doi.org/10.3989/graellsia.2011.v67.032>

Peters, J., Verhoest, N.E.C., Samson, R., Van Meirvenne, M., Cockx, L., De Baets, B., 2009. Uncertainty propagation in vegetation distribution models based on ensemble classiﬁers. Ecol. Model. 220 (6), 791–804, [http://dx.doi.org/10.1016/j.ecolmodel.](http://dx.doi.org/10.1016/j.ecolmodel)

2008.12.022

R Development Core Team, 2012. [R:](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [A](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [Language](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [and](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [Environment](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [for](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [Statistical](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320) [Computing.](http://refhub.elsevier.com/S0304-3800(15)00176-3/sbref0320)

Soares, S., Antunes, C.H., Araújo, R., 2013. Comparison of a genetic algorithm and simulated annealing for automatic neural network ensemble development. Neu- rocomputing 121, 498–511, <http://dx.doi.org/10.1016/j.neucom.2013.05.024>

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecol. Lett. 10 (12), 1115–1123, [http://dx.doi.org/10.1111/j.1461-0248.](http://dx.doi.org/10.1111/j.1461-0248) 2007.01107.x

Vezza, P., Mun˜ oz-Mas, R., Martinez-Capel, F., Mouton, A., 2015. Random forests to evaluate biotic interactions in ﬁsh distribution models. Environ. Model. Softw. 67, 173–183, <http://dx.doi.org/10.1016/j.envsoft.2015.01.005>

Wang, D., Alhamdoosh, M., 2013. Evolutionary extreme learning machine ensem- bles with size control. Neurocomputing 102, 98–110, <http://dx.doi.org/10.1016/> j.neucom.2011.12.046

Watts, M.J., Worner, S.P., 2008. Comparing ensemble and cascaded neural networks that combine biotic and abiotic variables to predict insect species distribution. Ecol. Inform. 3 (6), 354–366, <http://dx.doi.org/10.1016/j.ecoinf.2008.08.003>

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., et al., 2013. The role of biotic interactions in shaping distributions and realised assem- blages of species: implications for species distribution modelling. Biol. Rev. 88 (1), 15–30, <http://dx.doi.org/10.1111/j.1469-185X.2012.00235.x>

Yao, X., Xu, Y., 2006. Recent advances in evolutionary computation. J. Comput. Sci.

Technol. 21 (1), 1–18, <http://dx.doi.org/10.1007/s11390-006-0001-4>

Zhou, Z.H., Wu, J., Tang, W., 2002. Ensembling neural networks: many could be bet- ter than all. Artif. Intell. 137 (1–2), 239–263, <http://dx.doi.org/10.1016/s0004-> 3702(02)00190-x.