UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departament d'Enginyeria Hidràulica i Medi Ambient

Ph.D. Dissertation

Ecological models at fish community and species level to support effective river restoration

Presented by: Esther Julia Olaya Marín

Supervisor: Francisco Martínez Capel

March 2013

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departament d'Enginyeria Hidràulica i Medi Ambient

ECOLOGICAL MODELS AT FISH COMMUNITY AND SPECIES LEVEL TO SUPPORT EFFECTIVE RIVER RESTORATION

Presented by:
Esther Julia Olaya Marín

Supervisor:
Dr. Francisco Martínez Capel

Ph.D. Dissertation

Valencia, March 2013

The cover photograph shows the Cabriel River, Spain (photograph by Virginia Garófano-Gómez). Fish species at the lower part of the cover are *Squalius valentinus, Luciobarbus guiraonis*, *Cobitis paludica* (source: http://www.mediterranea.org/cae/divulgac/guipeces.htm), *Anguilla Anguilla* (source: Greenpeace, Spain), and *Parachondrostoma arrigonis* (source: Kalous and Doadrio).

This dissertation should be cited as:

Olaya-Marín, E.J. (2013). Ecological models at fish community and species level to support effective river restoration. PhD Thesis, Universitat Politècnica de València, Spain.

For Miguel, who offered me unconditional love and support throughout the course of this thesis.

ACKNOWLEDGMENTS

I am especially indebted to my supervisor Francisco Martínez Capel, for his guidance, encouragement and support in the development of this PhD thesis.

I am heartily thankful to Paolo Vezza, Juan Diego Alcaraz, and Rafael García Bartual for their interesting comments and suggestions to improve the research.

To all my wonderful friends, specially, Virginia Garófano, Paolo Vezza and Rafa Muñoz for their support and for being very good friends.

Thanks to everyone in the Department of Hydraulic Engineering and Environment of Universitat Politècnica de València, for their continued support.

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the projects SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and POTECOL "Evaluación del Potencial Ecológico de Ríos Regulados por Embalses y Desarrollo de Criterios para su mejora según la Directiva Marco del Agua" (CGL2007-66412).

I thank to Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment) and to the office of biodiversity affairs (Servicio de Biodiversidad, Conselleria de Medio Ambiente, Agua, Urbanismo y Vivienda, Comunidad Valenciana), who provided data to develop this study. Thanks to Javier Ferrer Polo and Amparo Piñón (Confederación Hidrográfica del Júcar), they partially supported data acquisition and discussion of the thesis. I am grateful to Juan Jiménez Pérez (Comunidad Valenciana) and Enrique Montero from Junta de Comunidades de Castilla-La Mancha, for the data provided to develop this study.

ABSTRACT

Native fish are indicators of the health of aquatic ecosystems, and they have become a key quality element to assess the ecological status of rivers. The understanding of factors affecting native fish species is important for the management and conservation of aquatic ecosystems. The general objective of this thesis are to analyse the relationships between biological and habitat variables (including connectivity) across a range of spatial scales in Mediterranean rivers, with the development of modelling tools to support the decision-making in river restoration.

This thesis is composed by four articles. The first aims to model the relationship between a set of environmental variables and native species richness (NFSR), and to evaluate the potential effectiveness of river restoration actions to improve NFSR in the Júcar river basin. In order to solve these questions, an artificial neural network (ANN) modelling approach was carried out, using the Levenberg–Marquardt learning algorithm in the model training phase. The partial derivatives method was applied to determine the relative importance of input environmental variables. According to the results, ANN model combined variables describing riparian quality, water quality, and physical habitat and helped to identify the primary drivers of the NFSR patterns in Mediterranean rivers. In the second part of the study, the model was used to evaluate the effectiveness of two restoration actions in the Júcar River: the removal of two abandoned weirs and the consequent increase in the proportion of riffles. These simulations indicated that richness increases with the augmentation of channel length without artificial barriers and riffle proportion, and demonstrated the utility of ANN as a powerful tool to support decisions in the management and ecological restoration of Mediterranean rivers.

The second paper aims to determine the relative importance of the two main factors controlling the reduction of native fish species richness (NFSR), i.e. the interactions between aquatic species, habitat (including river connectivity) and biological variables (including invasive species) in the Júcar, Cabriel and Turia rivers. To this end, three ANN models were analysed: the first one built only with

biological variables, the second one only built with habitat variables and the third one was made with the combination of both groups of variables. The results show that habitat variables are the most important drivers for the distribution of NFSR, and demonstrate the ecological relevance of the developed models. The findings of this study highlight the need to propose mitigation measures related to improve the habitat as a means to conserve and restore these Mediterranean rivers.

The third paper seeks to compare the reliability and ecological relevance of two predictive models of fish species richness, based on artificial neural networks (ANNs) and random forests (RF). The relevance of the selected input variables of each model was evaluated based on ecological knowledge and supported by other researches. Both models were developed using a *k*-fold cross validation procedure and their performance were evaluated by three metrics: the determination coefficient (R^2), the Mean Square Error (MSE) and the adjusted determination coefficient (R^2adj). According to the results, RF obtained the best performance in training. But, the cross-validation procedure revealed that both techniques gave similar results (R^2=68% for RF and R^2=66% for ANN). The comparison of different ML methods is very helpful for the critical analysis of the results obtained from the models.

The fourth paper has the following purpose: to evaluate the ability of ANN to identify local stress factors affecting density and presence/absence of *Luciobarbus guiraonis* in the Júcar river basin district. We used multilayer feed-forward artificial neural networks (ANN) to represent nonlinear relationships between *L. guiraonis* descriptors and biological and habitat variables. The models predictive power was evaluated based on the Kappa statistic (*k)*, the correctly classified instances (CCI), and the area under the curve (AUC) of a receiver operator characteristic (ROC) plots. According to the results, the presence/absence of *L. guiraonis* is well predicted by the ANN model (CCI= 87%, AUC= 0.85 and *k*= 0.66). The prediction of density was moderate (CCI = 62%, AUC=0.71 and *k*= 0.43). The most significant variables that described the presence/absence were: solar radiation, drainage area and proportion of exotic fish species with a relative importance of 27.8%, 24.53% and 13.60%, respectively. In the density model, the most important variables were coefficient of variation of mean annual flows with a relative importance of 50.5% and proportion of exotic fish species with 24.4%. The models provides important information about the relation of *L. guiraonis* with biotic and habitat variables, this new knowledge could be used to support future studies and practical decisions for the management and conservation of this species in the Júcar River Basin District.

RESUMEN

Los peces nativos son indicadores de la salud de los ecosistemas acuáticos, y se han convertido en un elemento de calidad clave para evaluar el estado ecológico de los ríos. La comprensión de los factores que afectan a las especies nativas de peces es importante para la gestión y conservación de los ecosistemas acuáticos. El objetivo general de esta tesis es analizar las relaciones entre variables biológicas y de hábitat (incluyendo la conectividad) a través de una variedad de escalas espaciales en los ríos Mediterráneos, con el desarrollo de herramientas de modelación para apoyar la toma de decisiones en la restauración de ríos.

Esta tesis se compone de cuatro artículos. El primero tiene como objetivos modelar la relación entre un conjunto de variables ambientales y la riqueza de especies nativas (NFSR), y evaluar la eficacia de potenciales acciones de restauración para mejorar la NFSR en la cuenca del río Júcar. Para ello se aplicó un enfoque de modelación de red neuronal artificial (ANN), utilizando en la fase de entrenamiento el algoritmo Levenberg-Marquardt. Se aplicó el método de las derivadas parciales para determinar la importancia relativa de las variables ambientales. Según los resultados, el modelo de ANN combina variables que describen la calidad de ribera, la calidad del agua y el hábitat físico, y ayudó a identificar los principales factores que condicionan el patrón de distribución de la NFSR en los ríos Mediterráneos. En la segunda parte del estudio, el modelo fue utilizado para evaluar la eficacia de dos acciones de restauración en el río Júcar: la eliminación de dos azudes abandonados, con el consiguiente incremento de la proporción de corrientes. Estas simulaciones indican que la riqueza aumenta con el incremento de la longitud libre de barreras artificiales y la proporción del mesohabitat de corriente, y demostró la utilidad de las ANN como una poderosa herramienta para apoyar la toma de decisiones en el manejo y restauración ecológica de los ríos Mediterráneos.

El segundo artículo tiene como objetivo determinar la importancia relativa de los dos principales factores que controlan la reducción de la riqueza de peces (NFSR), es decir, las variables del hábitat (incluyendo la conectividad fluvial) y las

biológicas (incluidas las especies invasoras) en los ríos Júcar, Cabriel y Turia. Con este fin, tres modelos de ANN fueron analizados: el primero fue construido solamente con variables biológicas, el segundo se construyó únicamente con variables de hábitat y el tercero con la combinación de estos dos grupos de variables. Los resultados muestran que las variables de hábitat son los "drivers" más importantes para la distribución de NFSR, y demuestran la importancia ecológica de los modelos desarrollados. Los resultados de este estudio destacan la necesidad de proponer medidas de mitigación relacionadas con la mejora del hábitat (incluyendo la variabilidad de caudales en el río) como medida para conservar y restaurar los ríos Mediterráneos.

El tercer artículo busca comparar la fiabilidad y relevancia ecológica de dos modelos predictivos de NFSR, basados en redes neuronales artificiales (ANN) y random forests (RF). La relevancia de las variables seleccionadas por cada modelo se evaluó a partir del conocimiento ecológico y apoyado por otras investigaciones. Los dos modelos fueron desarrollados utilizando validación cruzada *k*-fold y su desempeño fue evaluado a través de tres índices: el coeficiente de determinación (R^2), el error cuadrático medio (MSE) y el coeficiente de determinación ajustado (R^2_{adj}). Según los resultados, RF obtuvo el mejor desempeño en entrenamiento. Pero, el procedimiento de validación cruzada reveló que ambas técnicas generaron resultados similares (R^2 = 68% para RF y R^2 = 66% para ANN). La comparación de diferentes métodos de machine learning es muy útil para el análisis crítico de los resultados obtenidos a través de los modelos.

El cuarto artículo tiene como objetivo evaluar la capacidad de las ANN para identificar los factores que afectan a la densidad y la presencia/ausencia de *Luciobarbus guiraonis* en la demarcación hidrográfica del Júcar. Se utilizó una red neuronal artificial multicapa de tipo feed-forward (ANN) para representar relaciones no lineales entre descriptores de *L. guiraonis* con variables biológicas y de hábitat. El poder predictivo de los modelos se evaluó con base en el índice Kappa (k), la proporción de casos correctamente clasificados (CCI) y el área bajo la curva (AUC) característica operativa del receptor (ROC). La presencia/ausencia de *L. guiraonis* fue bien predicha por el modelo ANN (CCI = 87%, AUC = 0.85 y k = 0.66). La predicción de la densidad fue moderada (CCI = 62%, AUC = 0.71 y k

= 0.43). Las variables más importantes que describen la presencia/ausencia fueron: radiación solar, área de drenaje y la proporción de especies exóticas de peces con un peso relativo del 27.8%, 24.53% y 13.60% respectivamente. En el modelo de densidad, las variables más importantes fueron el coeficiente de variación de los caudales medios anuales con una importancia relativa del 50.5% y la proporción de especies exóticas de peces con el 24.4%. Los modelos proporcionan información importante acerca de la relación de *L. guiraonis* con variables bióticas y de hábitat, este nuevo conocimiento podría utilizarse para apoyar futuros estudios y para contribuir en la toma de decisiones para la conservación y manejo de especies en los en los ríos Júcar, Cabriel y Turia.

RESUM

Els peixos natius són indicadors de la salut dels ecosistemes aquàtics, i han esdevingut un element de qualitat clau per a avaluar l'estat ecològic dels rius. La comprensió dels factors que afecten a les espècies natives de peixos és important per a la gestió i conservació dels ecosistemes aquàtics. L'objectiu general d'aquesta tesi és analitzar les relacions entre variables biològiques i d'hàbitat (incloent la connectivitat) a través d'una varietat d'escales espacials als rius mediterranis, amb el desenvolupament d'eines de modelització per a donar suport a la presa de decisions en la restauració de rius.

Aquesta tesi es compon de quatre articles. El primer té com a objectius modelitzar la relació entre un conjunt de variables ambientals i la riquesa d'espècies natives (NFSR), i avaluar l'eficàcia d'accions potencials de restauració per a millorar l'NFSR a la conca del riu Xúquer. A fi de resoldre aquestes qüestions, es va aplicar un enfocament de modelització de xarxa neuronal artificial (ANN), i per a fer-ho es va utilitzar en la fase d'entrenament l'algorisme Levenberg-Marquardt. Es va aplicar el mètode de les derivades parcials per determinar la importància relativa de les variables ambientals. Segons els resultats, el model d'ANN combina variables que descriuen la qualitat de la ribera, la qualitat de l'aigua i l'hàbitat físic, i va ajudar a identificar els principals factors que condicionen el patró de distribució de l'NFSR als rius mediterranis. En la segona part de l'estudi es va utilitzar el model per avaluar l'eficàcia de dues accions de restauració al riu Xúquer: l'eliminació de dos assuts abandonats i l'increment consegüent de la proporció de corrent. Aquestes simulacions indiquen que la riquesa augmenta en incrementar la longitud lliure de barreres artificials i la proporció del mesohàbitat de corrent, i va demostrar la utilitat d'ANN com una eina poderosa per a donar suport a la presa de decisions referents a la gestió i la restauració ecològica dels rius mediterranis.

El segon article té com a objectiu determinar la importància relativa dels dos factors principals que controlen la reducció de la riquesa de peixos (NFSR), és a dir, les interaccions entre les espècies aquàtiques, variables de l'hàbitat (incloent-hi

la connectivitat fluvial) i biològiques (incloses les espècies invasores) als rius Xúquer, Cabriol i Túria. Amb aquest objectiu es va analitzar tres models d'ANN: el primer es va construir solament amb variables biològiques; el segon, únicament amb variables d'hàbitat; i el tercer, amb la combinació d'aquests dos grups de variables. Els resultats mostren que les variables d'hàbitat són els "drivers" més importants per a la distribució d'NFSR, i demostren la importància ecològica dels models desenvolupats. Els resultats d'aquest estudi destaquen la necessitat de proposar mesures de mitigació relacionades amb el millorament de l'hàbitat (incloent la variabilitat de cabals en el riu) com a mesura per a conservar i restaurar els rius mediterranis.

El tercer article cerca comparar la fiabilitat i rellevància ecològica de dos models predictius d'NFSR, basats en xarxes neuronals artificials (ANN) i random forests (RF). La rellevància de les variables seleccionades per cada model es va avaluar a partir del coneixement ecològic fonamentat en altres investigacions. Els dos models van ser desenvolupats utilitzant la validació creuada de *k* iteracions (*k*-fold) i el funcionament en va ser avaluat a través de tres índexs: el coeficient de determinació (R^2), l'error quadràtic mitjà (MSE) i el coeficient de determinació ajustat (R^2_{adj}). Segons els resultats, RF va tenir el millor funcionament en entrenament. Però el procediment de validació creuada va revelar que ambdues tècniques van generar resultats similars (R^2 = 68% per a RF i R^2 = 66% per a ANN). La comparació de diferents mètodes d'aprenentatge (machine learning) és molt útil per a l'anàlisi crítica dels resultats obtinguts a través dels models.

El quart article té com a objectiu avaluar la capacitat d'ANN per a identificar els factors d'estrès que afecten a la densitat i la presencia/absència de Luciobarbus guiraonis en la demarcació hidrogràfica del Xúquer. Es va utilitzar una xarxa neuronal artificial multicapa de tipus feed-forward (ANN) per a representar relacions no lineals entre descriptors d'L. guiraonis amb variables biològiques i d'hàbitat. El poder predictiu dels models es va avaluar sobre la base de l'índex Kappa (*k*), la proporció de casos correctament classificats (CCI) i l'àrea sota la corba (AUC) característica operativa del receptor (ROC). Segons els resultats, la presencia/absència d'*L. guiraonis* va ser ben predita pel model ANN (CCI = 87%, AUC = 0.85 i *k* = 0.66). La predicció de la densitat va ser moderada (CCI = 62%, AUC = 0.71 i k = 0.43). Les variables més importants que descriuen la presencia/absència van ser la radiació solar, l'àrea de drenatge i la proporció d'espècies exòtiques de peixos, amb un pes relatiu del 27.8%, el 24.53% i el 13.60% respectivament. En el model de densitat, les variables més importants van ser el coeficient de variació dels cabals mitjans anuals, amb una importància relativa del 50.5%, i la proporció d'espècies exòtiques de peixos, amb el 24.4%. Els models proporcionen informació important sobre la relació d'*L. guiraonis* amb variables biòtiques i d'hàbitat; aquest nou coneixement podria utilitzar-se per a fonamentar futurs estudis i per a contribuir a la presa de decisions per a la conservació i la gestió d'espècies als rius Xúquer, Cabriol i Túria.

CONTENTS

Chapter 1. Introduction………………………………………………. 23

Chapter 2. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain)………………………………………................................... 39

Chapter 3. Role of invasive species and habitat degradation on freshwater native fish diversity in Mediterranean River Basins.. 75

Chapter 4. A comparison of artificial neural networks and random forests to predict native fish species richness in Mediterranean rivers………………………..……………………………………… 109

Chapter 5. Modelling factors affecting the presence/absence and density of *Luciobarbus guiraonis* (Júcar River Basin, Spain)…….. 139

Chapter 6. Conclusion and future work…………………………..... 171

Chapter 1

Introduction

The populations of freshwater Mediterranean fish species are clearly decreasing as a consequence of habitat loss and degradation. Pollution, overexploitation of water resources, the introduction of invasive species and the alteration of riparian environments are the most important stressors which increase species risk extinction (Doadrio and Aldeguer, 2007; Smith and Darwall, 2006). The development of mathematical tools is very important to understand the effects of these alterations in the river ecological processes and advance on the knowledge about Mediterranean freshwater ecosystems. Knowledge improvement is necessary to design efficient and effective management and restoration measures (Cowx and Portocarrero Aya, 2011; Hurford *et al.,* 2010; Tirelli *et al.*, 2009).

Fish communities can be used as indicators of rivers environmental degradation, because they are very sensitive to water quality changes (Angermeier and Davideanu, 2004; Karr, 1981); thus, the study of fish ecology has been recognized as an important topic to assess the impact of human disturbances on aquatic ecosystems and evaluate the effect of management and restoration actions in water bodies (Bond and Lake, 2003; Cheng *et al.*, 2012; Lake *et al.*, 2007; Olaya-Marín *et al.,* 2012; Zarkami *et al.,* 2012). Currently, the prediction of species distributions and the characterization of the factors affecting species range limits are a challenge in freshwater ecology (Broennimann *et al.,* 2012). Therefore, modelling the descriptors of fish communities as a function of biotic and abiotic environmental predictors is a key issue to improve processes understanding in freshwater ecology

(Cheng *et al.,* 2012; Knudby *et al.,* 2010) and support decision-making in water management and conservation (Drew *et al.,* 2011; Jopp *et al.,* 2011; Lek *et al.,* 2005; Olaya-Marín *et al.,* 2012; Zarkami *et al.,* 2012). The application of machine learning techniques in ecological studies has been increased in recent years, due to its capacity to model complex and non-linear relationships, these techniques are not constrained by traditional assumptions about the statistical distribution of data (Gevrey *et al.,* 2004; Kohonen, 2001; Olden *et al.,* 2008).

Ecological models represent the interactions and changes of environmental elements and simulate the dynamics of spatial-temporal patterns of ecological processes (Drew *et al.,* 2011; Jopp *et al.,* 2011). This PhD thesis is focused on the development of predictive models for native fish as a function of physico-chemical, hydromorphological and biological variables to support the design of conservation and river restoration actions in Mediterranean rivers. For this aim, the models were built using artificial neural networks (ANN), which are considered in literature as powerful tools to address ecological data mining analysis (Brosse *et al.,* 2003; Franklin, 2010; Lek *et al.,* 2005; Olden *et al.,* 2008). The main advantages of ANNs are their high performance in the solution of non-linear problems, their ability to deal with noisy data (Goh, 1995; Olden *et al.,* 2008; Tirelli *et al.,* 2009), and their tolerance to the lack of independence, homoscedasticity and normality in datasets (Goh, 1995; Lek *et al.,* 2005; Olden *et al.,* 2008). ANNs can deal with the inherent variability of biological datasets, therefore this models are better to recognize patterns and make improved predictions than traditional statistical methods (Kang *et al.,* 2011).

Native fish richness is modelled in chapter 2, as an indicator of the effects of hydromorphological enhancements in river restoration. Chapter 3 addresses the role of invasive species and habitat degradation on freshwater native fish diversity in Mediterranean river basins. Chapter 4 presents a comparison of artificial neural networks and random forests modelling to predict native fish species richness in Mediterranean rivers. In the chapter 5, biological and habitat relationships are modelled to analyse the distribution of *Luciobarbus guiraonis* in the Júcar River Basin District. Finally, chapter 6 shows the general conclusions and future research questions derived from this PhD thesis.

The present dissertation includes the following research articles:

- Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernández JD. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). *Science of the Total Environment*. 2012; 440: 95-105.

- Olaya-Marín EJ, Martínez-Capel F, Vezza, P. A comparison of artificial neural networks and random forests to predict native fish species richness in Mediterranean rivers. *Knowledge and Management of Aquatic Ecosystems*. 2013; (under review).

1.1 CONTEXT

1.1.1 BACKGROUND

Machine learning (ML) has been an emerging discipline of Ecoinformatics during the last decade (Jorgessen *et al.,* 2009); its general objective is to analyse high complex and non-linear data structures, and making accurate predictive models (Drew *et al.,* 2011; Olden *et al.,* 2008). In recent years, ML has been considered as a powerful tool to model ecological data. This is explained by the fact that ecological data exhibit a variety of problems, such as complex data interactions and dependence between observations (Olden *et al.,* 2008; Olden *et al.,* 2006). A key advantage of ML is its better capacity to discover complex relationships and spatial patterns than traditional statistical models that assumes data normality (Jorgensen and Bendoricchio, 2001).

ML techniques can be classified in two categories; on one hand, *supervised learning* try to model the relationships between a set of inputs and a known output, such as multilayer perceptron artificial neural networks, random forests (Hogeweg, 1988), decision trees, bagging, boosting evolutionary algorithms, bayesian networks, nearest neighbors and support vector machines (Breiman, 2001). On the other hand, *unsupervised learning* reveals ecological data patterns based on input datasets. These models are known as self-organized systems, some of them are self-organizing maps (SOM), Hopfield networks, learning matrix, temporal associative memory (LAM), fuzzy associative memory, additive Grossberg, and adaptive resonance theory (Hopfield, 1982). Artificial neural networks (ANNs) and random forests (RF) have been successfully used by many researchers to design species distribution models and in other ecological issues (Drew *et al.*, 2011; Franklin, 2010; He *et al.*, 2010; Mouton *et al.*, 2011; Olaya-Marín *et al.*, 2012; Olden *et al.*, 2008; Recknagel, 2001). ANNs and RF are able to carry out complex computations in pattern recognition, signal classification and prediction problems. Both techniques can be used to successfully address complex modelling issues (Cutler *et al.*, 2007; Edia *et al.*, 2010; Prasad *et al.*, 2006; Tirelli and Pessani, 2009).

ANNs are mathematical models with an architecture inspired in the structure of the biological central nervous system. A biological neuron is constituted by three components: the cell body, dendrites and the axon (Fig 1a). Dendrites receive input signals, the cell body combines and integrate them and releases output signals. Axon transfers the information towards synaptic junctions, where it is distributed to new neuron sets. The fundamental building block of ANNs is the neuron (Chon and Park, 2006; Drew *et al.*, 2011; Franklin, 2010; Lek *et al.*, 2005) showed in Fig 1b. Dendrites are represented by input units; the weighted inputs are assimilated by the cell body, which acts as input attractor and nonlinear filter (activation function) to transfer the output signal to the axon (Jorgensen and Fath, 2011).

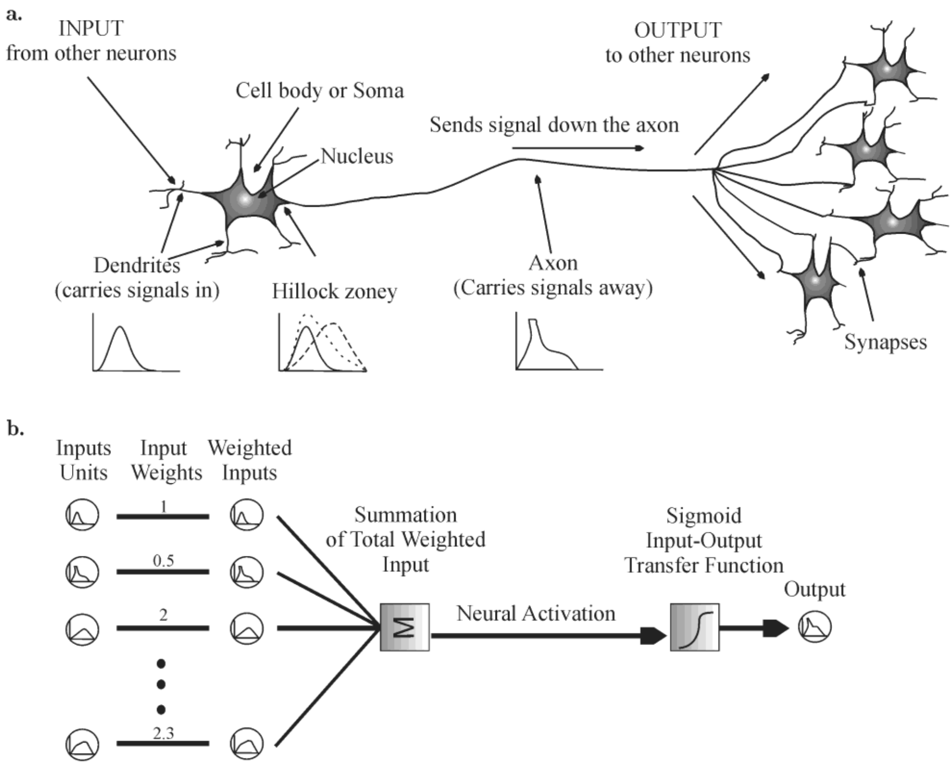

Fig. 1. Conceptual structures of biological and artificial neural networks. Modified from Jorgensen et al. (2009)

ANNs can be classified according to the neuron connections as: i) feed-forward neural network (signal propagates forward through layers), multilayer perceptron and radial basis functions are examples. ii) Recurrent networks (information flows feed-forward and feed-backward), some examples are recurrent multilayer perceptron, real-time recurrent network, and self-organizing maps. The most common type of ANN model is the multilayer perceptron (MLP) also called multilayer feed-forward neural networks (Fig. 2) and belong to supervised learning procedures.

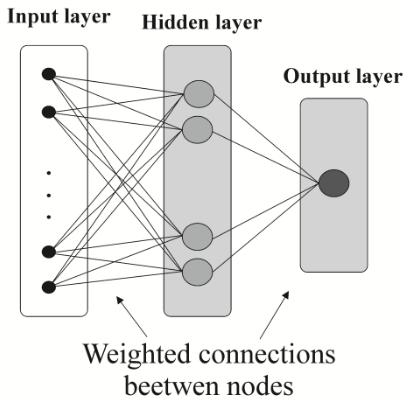

Fig. 2. Schematic diagram of tree-layered feed-forward artificial neural network (one input layer, one hidden layer and one output layer). Black circles are the input nodes of ANN model. These are linked to the hidden nodes (middle), which in turn are linked to an output node.

MLP is a layered feed-forward neural network, in which the neurons are organized in connected layers with unidirectional flow of information, from the input layer to the output one. A neuron in a layer is linked to all neurons in the adjacent layer, but there is neither connection within a layer nor feedback connections (Fig. 2). The number of input and output neurons depends on the number of predictors and target variables, respectively. The learning algorithms of MLP are based on a relatively simple concept (Fig. 3), the weights are changed in each iteration to minimize the error (represented by an objective function), the procedure ends when it is reached a stopping criteria. In a training phase, a set of input/output data is used to minimize the error function; once the optimum weights and bias are found the performance of the network is tested using an independent dataset.

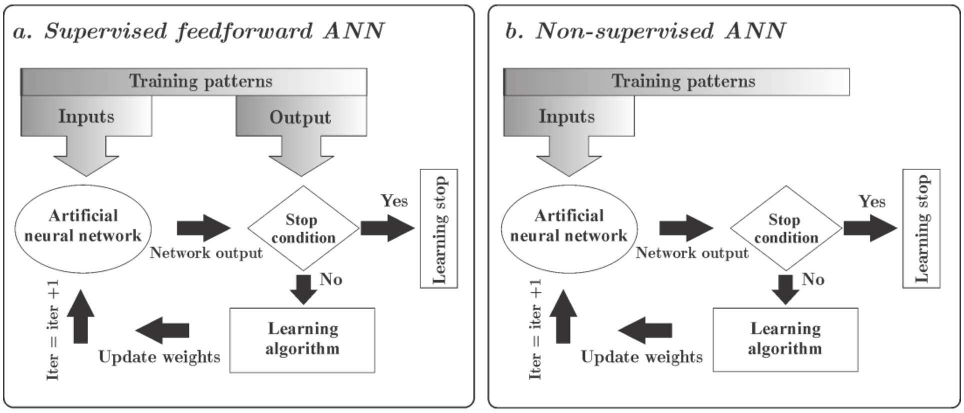

Fig. 3. Basic types of ANN. a) supervised feed-forward ANN; b) non-supervised ANN. Modified from Hernández (2006).

On the other hand, *early stopping* and *cross-validation* are the most commonly used criteria to avoid over-fitting in training phase, and some researchers have used them jointly (Sarle, 1995; Yong, 2006). *Early stopping* approach divides training dataset in two subsamples: The first one is the training sub-set, used to compute the gradient and update the weights and biases in the network. The second one is the validation sub-set. The error in the validation sub-set is observed through the training process, this error decreases in the first phase of training. However, when the optimization algorithm begins to over-fit the prediction to the training sub-set, validation error typically begins to rise and the training is stopped when the validation error increases for a specified number of iterations (Fig. 4), and the weights and biases at the minimum validation error are conserved as model parameters (Demuth *et al.*, 2010). This technique stops training before the network starts to learn from the noise present in training sub-set and prevents over-fitting improving the generalization capability of the trained neural network (Yong, 2006).

Once the model is trained, its performance is evaluated on an independent dataset (test dataset). This procedure is useful to objectively assess the level of likelihood of the prediction to a new input set (Demuth *et al.*, 2010) and learn about the generalization ability of the neural network.

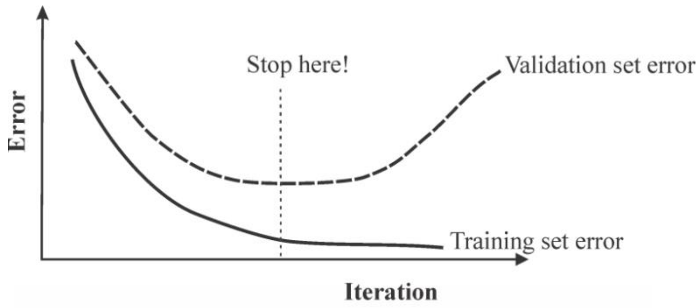

Fig.4. Neural network training process with early stopping criteria.

Cross-validation is employed in ecology when the size of dataset is insufficient to divide data in training and test subsets (Goethals *et al.,* 2007; Hastie *et al.,* 2009). The most used *cross-validation* method is *k*-fold; in which, dataset is divided in *k* subsets and the ANN model is trained with *k-1* sets, and validated with the another one (Goethals *et al.,* 2007). This training/test process is repeated *k* times, using different sub-sets in validation, this procedure is illustrated in Figure 5 using *k = 4* (Hastie *et al.,* 2009). Finally, model performance is the average of the error values in the *k* trials, and it is a global prediction error (Witten and Frank, 2005).

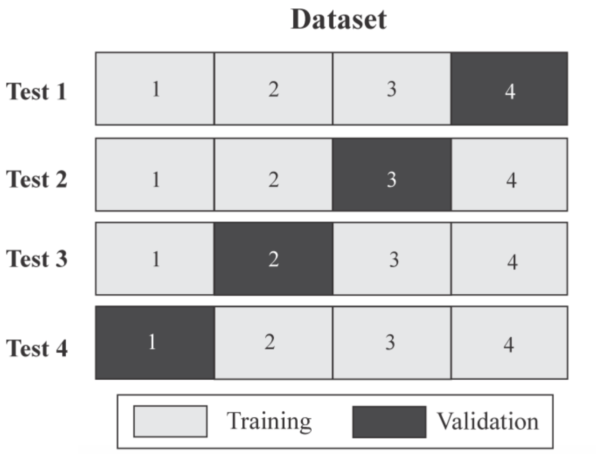

Figure 5. Scheme of the dataset division for a k-fold cross-validation using k = 4.

According to the aforementioned, ANN is a valuable technique to be applied in ecological modelling; its utility is focused on a better understanding of habitat and biological relationships, and the improvement of models predictability for the conservation and ecological restoration of rivers (Olden *et al.*, 2008; Tirelli *et al.*, 2009).

1.1.2 MOTIVATION

The European water framework directive (2000/60/EC) (WFD) establishes a framework for Community action in water management. Water is considered more than a resource; it is valued as a fundamental dimension of freshwater ecosystems and relevant to support a good environmental quality (Munné and Prat, 2004). The main aims of water conservation policy cover the management of the whole ecosystem; therefore, biological quality indicators are as important as physico-chemical quality indicators (CHE, 2007). Fish are sensitive to persistent environmental changes, which can be observed by the alteration of native fish species abundance and specific richness (CHJ, 2007; Laws, 2000); therefore, these indicators can be used to evaluate the health of river ecosystems and the effect of river restoration actions. For this reason, the WFD takes into account continental fish as one of the biological indicators to evaluate the ecological status of water bodies.

The development of mathematical models in ecology is fundamental to improve the scientific knowledge, and useful to support decision making from a practical point of view (Omlin and Reichert, 1999). Due to anthropogenic environment alterations, the utility of prediction and simulation in applied ecology have increased in recent years, with the aim of designing rehabilitation actions (Jopp *et al.*, 2011; Olden *et al.*, 2010). However, model predictability is constrained by ecosystems complexity, non-linearities of the emerging relationships (Olden *et al.*, 2008) and unknown mathematical structures governing the biological-habitat processes. According to Jopp *et al.* (2011), predictive models are useful for complementing existing approaches in at least five areas of research (Jopp *et al.*, 2011): a) Decision-support for conservation biology; b) Testing specific hypotheses, e.g. on the spatial scale of habitat selection; c) Generating hypotheses, e.g. on correlation of species traits with environmental variables, which can be tested experimentally; d) Identifying hierarchies of environmental drivers; and e) Prospective design of surveys, e.g. optimizing sampling schemes for rare species.

A recent avenue of approaches to improve predictability and learn about the behavior of freshwater ecosystems is based on ML techniques. Hence, ML is valuable to build mathematical models of species' distribution, abundance or diversity (Drew *et al.*, 2011; Guisan and Thuiller, 2005; He *et al.*, 2010; Olaya-Marín *et al.*, 2012; Zarkami *et al.*, 2012), and also to understand the biological-habitat processes to design conservation protocols and restoration measures, to assess the impact of human activities on natural resources, evaluate the impact of invasive species and study the impact of global warming on biodiversity and ecosystem (Franklin, 2010; Tirelli and Pessani, 2009; Zarkami *et al.*, 2012). There is no consensus in the scientific community about fundamental quantitative relationships to explain biological communities' descriptors as function of environmental-biological predictors through global, regional and local scales. Therefore, it is important to enrich the scientific literature with researches concerning mathematical models development and to provide tools to generate, integrate and synthesize knowledge in freshwater ecology through different climatic conditions around the world.

In the context of Mediterranean river basins, the development of new conceptualizations, theories and methods are needed to explain the relationships between biological and habitat variables and improve the understanding of ecological processes through a wide range of spatial and temporal scales. Special attention is required to native ichthyofauna, which is constituted by many endemic taxa with a very limited or local distribution. The freshwater fish fauna in Mediterranean river basins is particularly threatened, due to its high level of endemism. According to IUCN, the 56% of freshwater Mediterranean species are threatened (Smith and Darwall, 2006).

1.2 SCOPE AND RESEARCH PROBLEM

In Mediterranean rivers many stressors (pollution, introduction of exotic species, hydrological regime alteration and habitat loss) have caused native fish population decline and/or extinction (Smith and Darwall, 2006). Some studies in Mediterranean environments supports this affirmation; e.g. Costa *et al.* (2012) found that habitat degradation has reduced the complexity of a Mediterranean river in Spain and enhanced the declining of the endangered Júcar nase (*Parachondrostoma arrigonis*). Hermoso *et al.* (2011) showed that the proliferation of invasive species is as strong threat to the persistence of native assemblages in highly fluctuating environments. And Aparicio *et al.* (2000) arrived to the conclusion that water pollution and modifications of the habitat were the most important anthropogenic factors affecting the changes in fish community integrity in south-eastern Pyrenean watersheds, in the Iberian Peninsula.

The Mediterranean freshwater ecosystems require the development of ecological models to explain the complex relationships between the habitat features (including connectivity) and the aquatic biota at different spatial scales. Therefore, it is important to develop new scientific knowledge in Mediterranean fluvial ecology to delineate political, administrative and technical guidelines to address watershed and river restoration plans and support the integrated management of the Mediterranean river basins. This PhD thesis seeks to characterize and improve the understanding of the factors governing the decline of native fish species in Mediterranean rivers; therefore, this research aims to solve the following questions: ¿what are the most relevant environmental variables for the conservation of native fish in the Júcar, Cabriel and Turia rivers? ¿What role invasive species play in the declining of native fish in these rivers? And ¿What should be the most appropriate restoration actions for the Mediterranean rivers?

1.3 OBJECTIVES

The general objective of this PhD thesis is:

- To analyse the relationships between biological and habitat variables across a range of spatial scales in Mediterranean rivers, with the development of modelling tools to support the decision-making in river restoration.

The above objective implies the following specific objectives:

- To build a database of fish descriptors, biological and environmental variables at different spatial scales in rivers of the Júcar river Basin District.

- Specific objectives of chapter 2: (i) to model the relationship between a set of environmental variables (associated with different scales and ecosystem components) and the native species richness (NFSR); (ii) to assess the importance of the most relevant environmental variables to predict NFSR; and (iii) to evaluate the potential effectiveness of river restoration actions to improve NFSR in the Júcar river basin.

- Specific objectives of chapter 3: to determine the relative importance of the two main factors in the reduction of native fish species richness (NFSR), i.e. habitat (including water quality and river connectivity) and ecological interactions among aquatic species, in the Júcar, Cabriel and the Turia rivers.

- Specific objectives of chapter 4: (i) to compare the reliability and ecological relevance of two predictive models for fish richness based on the techniques of ANN and RF and (ii) to evaluate the concordance in terms of selected important variables between the two modelling approaches.

- Specific objectives of chapter 5: (i) to identify relevant environmental variables and model density and presence/absence of L. guiraonis and (ii) to assess the importance of each predictive environmental variable in the estimation of density and presence/absence.

1.4 REFERENCES

Angermeier PL, Davideanu G. Using fish communities to assess streams in Romania: Initial development of an Index of biotic integrity. Hydrobiologia 2004; 511: 65-78.

Aparicio E, Vargas MJ, Olmo JM, de Sostoa A. Decline of native freshwater fishes in a Mediterranean watershed on the Iberian Peninsula: A quantitative assessment. Environ. Biol. Fishes 2000; 59: 11-19.

Bond NR, Lake PS. Characterizing fish–habitat associations in streams as the first step in ecological restoration. Austral Ecology 2003; 28: 611-621.

Breiman L. Random Forests. Mach. Learn. 2001; 45: 5-32.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, *et al.* Measuring ecological niche overlap from occurrence and spatial environmental data. Glob. Ecol. Biogeogr. 2012; 21: 481-497.

Brosse S, Arbuckle CJ, Townsend CR. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. Biodivers. Conserv. 2003; 12: 2057-2075.

Costa RMS, Martínez-Capel F, Muñoz-Mas R, Alcaraz-Hernández JD, Garófano-Gómez V. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (river Cabriel, Spain). River Res. Appl. 2012; 28: 740-752.

Cowx IG, Portocarrero Aya M. Paradigm shifts in fish conservation: moving to the ecosystem services concept. J. Fish Biol. 2011; 79: 1663-1680.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, *et al*. Random Forests for classification in ecology. Ecology 2007; 88: 2783-2792.

CHE. Metodología para el establecimiento del estado ecológico según la Directiva Marco del Agua en la Confederación Hidrográfica del Ebro: Protocolos de muestreo y análisis para: fitoplancton, fitobentos(microalgas bentónicas), macrofitos, invertebrados bentónicos, ictiofauna. Madrid: Ministerio de Medio Ambiente; 2007.

Cheng L, Lek S, Lek-Ang S, Li Z. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 2012; 42: 127-136.

CHJ. La ictiofauna como elemento de calidad de los ríos de la demarcación hidrográfica del río Júcar. Valencia: Confederación Hidrográfica del Júcar; 2007.

Chon T-S, Park Y-S. Ecological informatics as an advanced interdisciplinary interpretation of ecosystems. Ecol. Inform. 2006; 1: 213-217.

Demuth H, Beale M, Hagan M. Neural Networks toolbox 6. Users Guide. Matlab. Natick, Massachusetts The MathWorks, Inc.; 2010.

Doadrio I, Aldeguer M. La invasión de especies exóticas en los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Drew CA, Wiersma Y, Huettmann F. Predictive species and habitat modeling in landscape ecology: concepts and applications. New York: Springer; 2011.
Edia EO, Gevrey M, Ouattara A, Brosse S, Gourène G, Lek S. Patterning and predicting aquatic insect richness in four West-African coastal rivers using artificial neural networks. Knowl. Managt. Aquatic Ecosyst. 2010.
Franklin J. Mapping species distributions: spatial inference and prediction. New York: Cambridge University Press; 2010.
Gevrey M, Rimet F, Park YS, Giraudel J-L, Ector L, Lek S. Water quality assessment using diatom assemblages and advanced modelling techniques. Freshw. Biol. 2004; 49: 208-220.
Goethals P, Dedecker A, Gabriels W, Lek S, De Pauw N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 2007; 41: 491-508.
Goh ATC. Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering 1995; 9: 143-151.
Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecology Letters 2005; 8: 993-1009.
Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference and Prediction: Springer; 2009.
He Y, Wang J, Lek-Ang S, Lek S. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. Sci. Total Environ. 2010; 408: 4211-4220.
Hermoso V, Clavero M, Blanco-Garrido F, Prenda J. Invasive species and habitat degradation in Iberian streams: an analysis of their role in freshwater fish diversity loss. Ecol. Appl. 2011; 21: 175-188.
Hernández L. Predicciones y optimización de emisores y consumo mediante redes neuronales en motores disel Universidad Politécnica de Valencia: Reverté S.A; 2006.
Hogeweg P. Cellular automata as a paradigm for ecological modeling. Appl. Math. Comput. 1988; 27: 81-100.
Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 1982; 79: 2554-2558.
Hurford C, Schneider M, Cowx I, editors. Conservation Monitoring in Freshwater Habitats: A Practical Guide and Case Studies. Dordrecht, Netherlands: Springer; 2010.
Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, Teachers and Scientists. Berlin: Springer-Verlag; 2011.
Jorgensen SE, Bendoricchio G, editors. Fundamentals of Ecological Modelling. In: ed r, editor. Developments in Environmental Modelling, 21. Oxford: Elsevier Science; 2001.

Jorgensen SE, Fath BD. Fundamentals of ecological modelling: applications in environmental management and research. 4th ed. Amsterdam: Elsevier; 2011.

Jorgessen S, E, Chon T-S, Recknagel F. Ecological Modelling and Informatics. Great Britain: WIT Press; Har/Com edition; 2009.

Kang HY, Rule RA, Noble PA. 9.09 - Artificial Neural Network Modeling of Phytoplankton Blooms and its Application to Sampling Sites within the Same Estuary. In: Editors-in-Chief: Eric W, Donald M, editors. Treatise on Estuarine and Coastal Science. Academic Press, Waltham, 2011, pp. 161-172.

Karr JR. Assessment of Biotic Integrity Using Fish Communities. Fisheries 1981; 6: 21-27.

Knudby A, Brenning A, LeDrew E. New approaches to modelling fish-habitat relationships. Ecol. Model. 2010; 221: 503-511.

Kohonen T. Self-Organizing Maps Berlin and New York: Springer-Verlag; 2001.

Lake PS, Bond N, Reich P. Linking ecological theory with stream restoration. Freshw. Biol. 2007; 52: 597-615.

Laws EA. Aquatic Pollution: An Introductory Text. United States of America: John Wiley & Sons; 2000.

Lek S, Scardi M, Verdonschot P, Descy JP, Park YS, editors. Modelling community structure in freshwater ecosystems. Berlin: Springer-Verlag; 2005.

Mouton AM, Alcaraz-Hernández JD, De Baets B, Goethals PLM, Martínez-Capel F. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Modell. Softw. 2011; 26: 615-622.

Munné A, Prat N. La diagnosis y mejor de los ecosistemas fluviales mediante la Directiva Marco del Agua. http://www.unizar.es/fnca/index3.php?id=1&pag=5 Fundación Nueva Cultura del Agua, 2004.

Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernández JD. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). Sci. Total Environ. 2012; 440: 95-105.

Olden JD, Kennard MJ, Leprieur F, Tedesco PA, Winemiller KO, García-Berthou E. Conservation biogeography of freshwater fishes: recent progress and future challenges. Divers. Distrib. 2010; 16: 496-513.

Olden JD, Lawler JJ, Poff NL. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Olden JD, Poff NL, Bledsoe BP. Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. Ecol. Inform. 2006; 1: 33-42.

Omlin M, Reichert P. A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Model. 1999; 115: 45-59.

Prasad A, Iverson L, Liaw A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 2006; 9: 181-199.

Recknagel F. Applications of machine learning to ecological modelling. Ecol. Model. 2001; 146: 303-310.

Sarle W. Stopped training and other remedies for overfitting, 1995.

Smith KG, Darwall WRT, editors. The status and distribution of freshwater fish endemic to the mediterranean basin. Gland, Switzerland/Cambridge, UK.: IUCN -The World Conservation Union; 2006.

Tirelli T, Pessani D. Use of decision tree and artificial neural network approaches to model presence/absence of *Telestes muticellus* in piedmont (North-Western Italy). River Res. Appl. 2009; 25: 1001-1012.

Tirelli T, Pozzi L, Pessani D. Use of different approaches to model presence/absence of *Salmo marmoratus* in Piedmont (Northwestern Italy). Ecol. Inform. 2009; 4: 234-242.

Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed. San Francisco: Morgan Kaufmann Publishers Inc.; 2005.

Yong L. Create Stable Neural Networks by Cross-Validation. Neural Networks, 2006. IJCNN '06. International Joint Conference on; 2006. p. 3925-3928.

Zarkami R, Sadeghi R, Goethals P. Use of fish distribution modelling for river management. Ecol. Model. 2012; 230: 44-49.

Chapter 2

Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain)

Esther Julia Olaya-Marín, Francisco Martínez-Capel, Rui Manuel Soares Costa, Juan Diego Alcaraz-Hernández

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

ABSTRACT

The richness of native fish is considered to be an indicator of aquatic ecosystem health, and improving richness is a key goal in the management of river ecosystems. An artificial neural network (ANN) model based on field data from 90 sample sites distributed throughout the Júcar River Basin District was developed to predict the native fish species richness (NFSR). The Levenberg-Marquardt learning algorithm was used for model training. When constructing the model, we tried different numbers of neurons (hidden layers), compared different transfer functions, and tried different *k* values (from 3 to 10) in the *k*-fold cross-validation method. This process and the final selection of key variables with relevant ecological meaning support the reliability and robustness of the final ANN model. The partial derivatives method was applied to determine the relative importance of input environmental variables. The final ANN model combined variables describing riparian quality, water quality, and physical habitat and helped to identify the primary drivers of the NFSR patterns in Mediterranean rivers. In the second part of the study, the model was used to evaluate the effectiveness of two restoration actions in the Júcar River: the removal of two abandoned weirs and the progressive increase in the proportion of riffles. The model indicated that the combination of these actions produced a rise in NFSR, which ultimately reached the maximum values observed in the reference site of that river ecotype (*sensu* the European Water Framework Directive). The results demonstrate the importance of longitudinal connectivity and riffle proportion for improving NFSR and the power of ANNs to help decisions in the management and ecological restoration of Mediterranean rivers. Furthermore, this model at the basin scale is the first step for further research on the effects of water scarcity and global change on Mediterranean fish communities.

Keywords: Artificial neural networks; River connectivity; Mitigation measures; Hydromorphology; Fish richness, River restoration.

2.1 INTRODUCTION

In 2000, the European Water Framework Directive (hereafter WFD) acknowledged the importance of freshwater fish communities as indicators that can be used to assess the ecological status of rivers (European Commission, 2000). Freshwater fish are considered to be good indicators of water quality in river systems (Angermeier and Davideanu, 2004; Karr, 1981) due their sensitivity to human disturbances, which alter community parameters such as the abundance of native species and species richness (Laws, 2000; Oberdoff *et al.,* 1995; Rosenberg and Resh, 1993). Species richness is a primary indicator used in the conservation and management of fish communities; it can be used as an indicator of ecological changes and as a criterion for the selection of conservation areas (He *et al.,* 2010; Lek *et al.*, 2005; van Jaarsveld *et al.,* 1998).

From the perspective of species conservation, the Mediterranean part of Europe has been recognised as a global biodiversity hotspot for freshwater fish species and for plant and terrestrial animal species (Cuttelod *et al.,* 2008). However, an ongoing extinction crisis is affecting Europe's freshwater fishes, and ambitious conservation actions, including the adequate protection and management of key freshwater habitats, are urgently needed (Freyhof and Brooks, 2011). The increased frequency and intensity of droughts are already impacting freshwater systems and the species that rely on them, especially in the Mediterranean region (Freyhof and Brooks, 2011). Conservation of fish diversity is one of the most critical issues facing the preservation of European biodiversity (Zitek *et al.,* 2008). Therefore, implementation of effective river restoration schemes is very important, and knowledge about the relationships between hydromorphological features and fish populations is essential for the design of effective actions.

In the last decade, predictive modelling of species distribution has become a powerful tool to support decisions in conservation and natural resource management (Drew *et al.,* 2011; Jopp *et al.,* 2011). Several studies have used conventional multivariate statistical methods to determine the effect of human disturbance on native species richness. For example, in the Iberian Peninsula, Corbacho and Sánchez (2001) analysed the factors affecting the native species richness in the Guadiana Basin (Spain) using principal component analysis and multiple regression analysis. Clavero *et al.* (2004) studied the effect of dams and introduced species on fish biodiversity in Iberian basins using generalised linear models. In relation to habitat connectivity, Alexandre and Almeida (2010) evaluated the effect of small artificial barriers on fish richness in the Muge and Erra Rivers using canonical correlation analysis and analysis of covariance.

Simultaneously, the development of advanced techniques in the machine learning area has allowed the creation of predictive models with greater power for explaining and predicting ecological patterns (Olden *et al.*, 2008); such models have the ability to model complex, nonlinear relationships in ecological data without having to satisfy the restrictive assumptions required by conventional, parametric approaches (Elith *et al.*, 2006; Olden and Jackson, 2002a; Recknagel, 2003). Examples of Machine Learning techniques are artificial neural networks (ANNs), random forests (RFs), genetic algorithms and support vector machines. ANNs have been used frequently and successfully in freshwater fish studies, giving satisfactory results in regards to learning capacity, adaptation, parallelization, speed, and flexibility (Soria *et al.*, 2010) and showing high efficiency in linking environmental variables, which are highly complex and nonlinear (Lek *et al.*, 2005).

However, only a few studies have used Machine Learning techniques to predict native fish species richness (NFSR). For example, Mastrorillo *et al.* (1998) found that the ANN is a powerful tool of prediction compared to traditional modelling methods. He *et al.* (2010) used RF and Classification and Regression Trees (CART) to predict endemic fish assemblages and species richness in the upper Yangtze River (China). Knudby *et al.* (2010) predicted fish species richness, diversity, and biomass in the reefs around the Chumbe and Bawe Islands (Tanzania) using different techniques (e.g., RFs, boosted regression trees, and support vector machines). Recently, Cheng *et al.* (2012) used RF and CART to predict species richness in lakes in the Yangtze River Basin.

These techniques also can be used to develop ecological response models. For example, studies at the fluvial network or basin scale that relate the biota with environmental variables allow the development of ecological response models with potential application in environmental flow assessments at the basin scale. These models can be applied to different environmental flow methodologies (see Paredes-Arquiola *et al.*, 2011 in press; Poff *et al.*, 2010), assuming that there is an economical way to up-scale instream flow studies in regions where there is a shortage of ecological species information (Paredes-Arquiola *et al.*, 2011 in press). Ecological response models can be used to complement and compare results from reach-scale assessments when it is necessary to extrapolate at the basin scale according to river types.

Understanding the relationships among hydrology, habitats, and fish populations is key to designing effective river restoration actions. In this work we used variables related to the physicochemical properties of water, hydromorphology (river flow, habitats), geographic location, and biological indexes of water and riparian quality to model the NFSR using an ANN and to simulate restoration actions in the main stem of the Júcar, Cabriel, and Turia Rivers. The objectives of this study were to: (1) model the relationship between a set of environmental variables (associated with different scales and ecosystem components) and the NFSR using the ANN; (2) assess the importance of the most relevant environmental variables for the NFSR; and (3) evaluate the potential effectiveness of a river restoration action (i.e., improving river connectivity) to improve NFSR in the Júcar River. The ANN model at the basin scale developed in this study is the first step in developing more complex simulations at smaller time scales (e.g., using daily flow data) and in assessing the effects of water scarcity and global change on Mediterranean fish communities.

2.2 MATERIALS AND METHODS

2.2.1 STUDY AREA AND DATA COLLECTION

The study area consists of the main stem of the Júcar, Cabriel, and Turia Rivers (Fig. 1), in the Júcar River Basin District (Eastern Iberian Peninsula). The three watersheds have a Mediterranean climate, and their environmental characteristics show a similar pattern of variability. The coldest and rainy areas are located in the mountainous zones, and the most temperate areas are situated near the coast (Estrela *et al.*, 2004). In Spain, the Ministry of Environment developed an official ecotype classification to implement the WFD (CEDEX, 2005). The upper reaches of the rivers have a small percentage of flow regulation (in relation to mean flow), and they belong to the Mediterranean calcareous mountain rivers (CMM) ecotype.

The large dams produce significant regulation downstream (Alarcón, Contreras, Benagéber, see Fig. 1). Downstream of the Alarcón and Contreras Dams lay segments of the mineralized Mediterranean-continental rivers (MCM) ecotype. The sampling sites located in the lowest reaches in the Turia River belong to the mineralized rivers of Mediterranean low mountain (ML) and low-altitude Mediterranean rivers (MML) ecotypes. In general, the flow regimes match the rain pattern and exhibit strong seasonal and annual variability. Consequently, severe droughts occur in summer and flash floods occur in winter and spring (Belmar *et al.,* 2010; Gasith and Resh, 1999; Vidal-Abarca *et al.,* 1992). The predominant lithographic groups are calcarenites and marls, although significant proportions of limestone and alluvial material are present. The forests cover a great percentage of the western mountainous areas, but the watersheds are highly anthropic in the eastern area (CHJ, 2007).

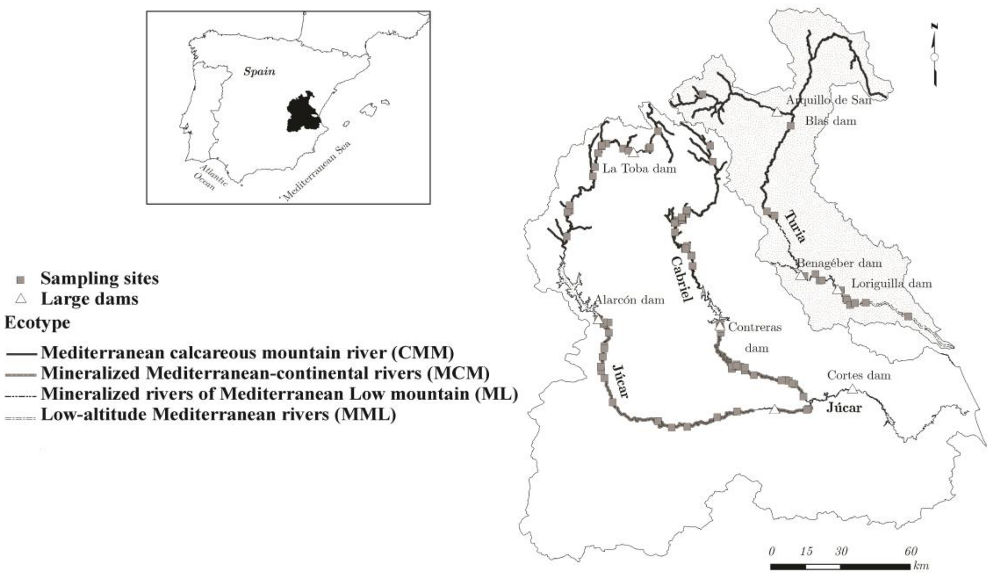

Fig. 1. Location of the Júcar, Cabriel and Turia River Basins in the Iberian Peninsula, with detail of the 90 sampling sites (square), the large dams (triangle) and the river's ecotypes (CEDEX, 2005) sensu the European Water Framework Directive.

The ecological importance of the fish communities in these rivers resides in their adaptation to Mediterranean conditions and to the number of endemic and endangered species. The communities are dominated by cyprinids (Ferreira *et al.*, 2007) and are characterised by a high number of endemic species, which have a reduced distribution range compared with other fish species elsewhere in Europe (Doadrio, 2001; Granado-Lorencio, 1996). However, information about habitat suitability for the fish communities is scarce, likely because of the low commercial fishing value of these species. In the reference sites for the assessment of ecological status (*sensu* the WFD), the highest NFSR values were as follows: In the reference site of the CMM ecotype in the Júcar River (NFSR = 5), the species present are *Salmo trutta fario*, *Luciobarbus guiraonis*, *Achondrostoma arcasii*, *Iberocypris alburnoides*, and *Cobitis paludica*. The reference site of the same ecotype in the Cabriel River (NFSR = 5) includes *S. trutta fario*, *Squalius pyrenaicus*, *L. guiraonis*, *A. arcasii*, and *Parachondrostoma arrigonis* (Júcar nase). In this location, the last is a very important species that is critically endangered (Freyhof and Brooks, 2011), and the only sustainable populations of the Júcar nase live in the Cabriel River and the Magro River (i.e., tributaries of the Júcar River) (Costa *et al.*, 2012). The Cabriel River is the only river where habitat suitability studies for the Júcar nase could be conducted (Costa *et al.*, 2012), and biological and ecological data about this species are scarce. In the same ecotype in the river Turia, the maximum NFSR is 3, and the species present are *L. guiraonis*, *Barbus haasi*, and *Parachondrostoma turiense*. The reference site of the MCM ecotype is located in the Cabriel River near Puente Tamayo (NFSR = 5), where the native fish species are *L. guiraonis*, *Salaria fluviatilis*, *S. pyrenaicus*, *Anguilla anguilla*, and *P. arrigonis* (the last two are critically endangered). Regarding the other ecotypes in the Turia River, the maximum NFSR is 4. In the ML ecotype, the species are *S. trutta fario*, *L. guiraonis*, *Squalius valentinus*, and *C. paludica*; in some reaches *S. trutta fario* is substituted by *P. turiense* or *A. anguilla*. In the same river but in the MML ecotype, *L. guiraonis*, *S. valentinus*, *P. turiense*, and *A. anguilla* are present.

2.2.2 RESPONSE VARIABLE AND PREDICTOR VARIABLES

In this study we used 90 fish sampling sites distributed along the three rivers (Fig. 1). Fish were captured by electrofishing (single-pass) that covered more than 50 m of river length. Sampling was conducted in the spring and summer months from 2005 to 2009. Each individual fish was identified to species.

Twenty-four environmental variables were used in the development of the ANN models (Table 1). In agreement with other investigations (Granado-Lorencio, 1996; Ibarra *et al.,* 2003; Jackson *et al.,* 2001; Lek *et al.,* 1996; Mastrorillo *et al.,* 1998), they were chosen according to their degree of importance for fish life and according to their availability in public databases. The environmental variables correspond to three groups: physicochemical parameters of water quality, hydromorphology, and biological indicators of water and riparian quality. We took into account different spatial scales, from mesohabitat measurements to larger scales such as the fluvial segment and the segment watershed (the latter was measured using a geographic information system (GIS)) (Table 1). As some studies have indicated, models based on multiple spatial scales usually outperform single-scale analyses (Olden *et al.*, 2006).

Water quality is widely recognised as a key factor affecting fish species distribution, as water pollution severely alters fluvial dynamics and compromises the survival of fish fauna (Granado-Lorencio, 2000; Jackson *et al.*, 2001). For this reason, physicochemical water variables and biological indicators of water and riparian quality were used in this study. The mesohabitats or hydromorphological units determine the available habitat for fish communities at the meso- and microhabitat-scale; thus, mesoscale variables are important for fish (Bernardo *et al.,* 2003; Costa *et al.,* 2012). The magnitude and variability of river discharge determine the lifecycle traits of Mediterranean fish fauna (Ferreira *et al.*, 2007; Granado-Lorencio, 2000). Other geographic variables, such as watershed area, distance from the headwater source, and altitude, affect fish species richness distribution patterns (Oberdorff *et al.*, 1995; Reyjol *et al.*, 2007). In this study, channel length without artificial barriers was taken into account because transverse hydraulic structures fragment fluvial continuity: act as barriers to reproductive fish migration, inhibit colonisation of empty reaches, and favour the development of lentic habitats, which are suitable for exotic species (Granado-Lorencio, 2000). Riverine vegetation plays an active role in the preservation of aquatic life (Naiman *et al.,* 1993; Patten, 1998), as it provides refuge and food and constitutes a biological buffer, which decreases the input of pollutants from alluvial and colluvial soils. Moreover, riverine vegetation controls the flood regime and water channel temperature (Hattermann *et al.,* 2006; Quinn *et al.,* 2004).

The physical and chemical properties of water were averaged for the year when fish were sampled at each of the 90 sampling points, using data recorded every 3 months as part of the official monitoring network of the Júcar River Basin Authority. The same authority provided the daily flow data from gauging stations. The mean monthly flows were estimated at each of the sampling points where there was no gauging station; the flow rate was interpolated between gauged sites using the relationship between the flow rate in the natural regime and the drainage area accumulated at the point (Caissie, 2006a; Caissie and El-Jabi, 1995; Leopold and Maddock, 1953; Leopold *et al.,* 1964). The mean annual flow in natural conditions was obtained by applying the water balance equation and the principle of mass conservation (Wurbs, 2006). The mean width of the water surface and the proportions of hydromorphological units were determined in the field based on the classification by Dolloff *et al.* (1993) and adapted to these Mediterranean rivers as pool, glide, riffle, run, and rapid. The same classification was applied in previous studies of Iberian rivers (Alcaraz-Hernández, 2011; Costa *et al.,* 2012). The geographical variables were determined using the ArcGIS™ 9.3.1 software (ESRI ©2009), based on the layers of the official river network supplied by the Júcar River Basin Authority. We used two biological indicators: the index of water quality based on aquatic invertebrates (IBMWP), (Alba-Tercedor and Sánchez-Ortega, 1988) and the index of riparian quality (Munné *et al.,* 2003). The second indicator considers four components: total riparian vegetation cover, cover structure, cover quality, and channel alterations. The IBMWP was obtained from field data (biomonitoring network) collected by the Júcar River Basin Authority. The riparian forest quality index (QBR) was determined in a basin-scale study funded by the same authority (Aguilella *et al.,* 2005).

Table 1. Potential predictive variables considered in the development of the model for native fish species richness (n = 90). For each variable, it is noted whether it was measured in situ or in a GIS; other data were supplied by the Júcar River Basin Authority (water quality and flow monitoring network, MN, or biomonitoring network, BMN).

Variable	Method	Code	Unit	Mean	Range
Physicochemical conditions of water					
Dissolved oxygen	MN	DIS	mg/l	9.5	8-11
Biological Oxygen Demand	MN	BOD	mg/l	2.5	2.0-4.0
Total phosphorus	MN	TOP	mg/l	0.06	0.02-0.22
Nitrites	MN	NIT	mg/l	0.02	0.01-0.23
pH	MN	PH	-	8.1	7.7-8.3
Suspended solids	MN	SUS	mg/l	11.3	3.1-25.2
Water conductivity	MN	CON	µS/cm	797.8	499.1-1210
Water temperature	MN	WAT	ºC	13.3	5.7-16.7
Hydromorphology					
Hydromorphological units:					
Pools (%)	*In situ*	POO	-	48.6	0.0-95.0
Glide (%)	*In situ*	GLI	-	11.2	0.0-80.0
Riffle (%)	*In situ*	RIF	-	28.2	0.0-89.0
Rapid (%)	*In situ*	RAP	-	5.7	0.0-53.0
Run (%)	*In situ*	RUN	-	6.0	0.0-50.0
Mean width of water surface	*In situ*	WID	m	12.4	3.1-19
Channel length without artificial barriers	GIS	CWB	km	26.3	0.5-95.8
Altitude	GIS	ALT	m a.s.l.	746.4	92.0-1363
Drainage area	GIS	DRA	km^2	3318.8	54.0-10952
Distance from headwater source	GIS	DHS	km	150.1	20.5-327.4
Mean Annual flow rate	MN	FMA	m^3/s	4.3	0.03-11.31
Inter-annual mean flow (calculated for 5 years)	MN	FIA	m^3/s	5.50	0.11-12.36
Coefficient of variation of mean monthly flows (fish sampling year)	MN	FIM	-	0.5	0.28-0.94
Coefficient of variation of mean annual flows (calculated for 5 years)	MN	FCV	-	0.4	0.15-0.81
Biological indices of water quality and riparian quality					
Iberian Biological Monitoring Working Party-IBMWP	BMN	IBMWP	-	131.6	64-260
Index of Riparian Habitat Quality	BMN	QBR	-	73.6	10-100

2.2.3 ARTIFICIAL NEURAL NETWORKS MODELLING

ANNs are mathematical models that are inspired by the structure and function of biological nervous systems (Olden *et al.,* 2008). In this study we built a multilayer perceptron neural network (MLP), as they are very frequently used for supervised learning and are the most used networks in ecology (Özesmi *et al.,* 2006). The learning procedure for this kind of neural network involves optimization of the mean squared error for a dataset of the output variable based on a set of predictive environmental variables. At each iteration, the weights or parameters are changed in order to find a minimum mean squared error. A detailed description of the MLP can be found in Dedecker *et al.* (2005), Goethals *et al.* (2007), Olden *et al.* (2008) and Mouton *et al.* (2010).

The first step in the model construction was the calculation of a correlation matrix to estimate the potential effect of collinearity and we used a cluster representation to visualise collinearity (Fig. 3). When two variables were highly correlated (Spearman's $rho^2 > 0.5$) we used the one with higher ecological interpretability (Brosse *et al.,* 1999; Dormann, 2011; van Wijk and Bouten, 1999). We then used the forward stepwise method to incorporate predictive variables into the network (Gevrey *et al.,* 2003). Many candidate MLP models were built to assess the optimal number of neurons in the hidden layer and the proper transfer function for the hidden and output layers. An important issue in ANN architecture is the selection of the transfer functions. Frequently, the neurons of a layer share the same kind of transfer function and the most used functions are the sigmoidal ones (Goethals *et al.,* 2007). However, the advantage of selecting a particular transfer function has not been demonstrated mathematically yet (Hassoun, 1995). Currently, the criterion for selecting a transfer function is the better performance of the model, as tested by trial and error (Isa *et al.,* 2010) .Therefore, we tried two combinations of transfer functions in the hidden and output layers (hidden/output): (1) hyperbolic tangent/lineal and (2) logistic/logistic. Data were scaled to ensure that predictive variables get equal attention during the training process (Maier and Dandy, 2000) and to commensurate their values with the limits (hyperbolic tangent [–1,+1] or logistic [0,+1]) of the transfer function used (Olden and Jackson, 2002b). The models were built with a single hidden layer; this is satisfactory for statistical applications (Bishop, 1996) and it notoriously decreases the computation time. Commonly, the use of one hidden layer yields similar results to the use multiple hidden layers (Kurková, 1992).

The Levenberg-Marquardt algorithm was used to train the neural networks; this is the fastest procedure to train neural networks of moderate size (Karul *et al.*, 2000) and it was recommended in previous studies (Gutiérrez-Estrada and Bilton, 2010; Tan and Van Cauwenberghe, 1999). To evaluate the predictive performance in the validation we used the method of *k*-fold cross-validation. This method is frequently used in ecology when the number of observations is not sufficient to divide the data into training and validation sets (Olden *et al.*, 2008). Goethals *et al.* (2007) suggested that it is necessary to test several values of *k* between 3 and 10. In this work, *k* was empirically determined by comparison of performance among the networks constructed with a range of *k* values between 3 and 10. A detailed description of this algorithm can be found in Shepherd (1997). Model performance was measured by the correlation coefficient (*r*) and the mean squared error (MSE).

A sensitivity analysis was performed to evaluate the contribution of each predictive variable to the ANN output. For this purpose the partial derivatives method (PaD) was applied (Dimopoulos *et al.*, 1999; Dimopoulos *et al.*, 1995). This method estimates the relative importance of each input variable to the prediction of NFSR in the model; a variable was considered relevant when its importance value was higher than 15% (Brosse *et al.*, 2003).

2.2.4 SIMULATION OF MITIGATION MEASURES

The ANN model with optimal performance was used to simulate the effect of restoration measures by introducing changes in the key predictive variables. Longitudinal fluvial connectivity and riffles are highly relevant to Iberian endemic cyprinid fishes (Granado-Lorencio, 1996; Ilhéu *et al.*, 1999), and mitigation measures related to these factors were simulated by changing the values of the following two variables: riffle proportion to river length (RIF) and channel length without artificial barriers (CWB). The removal of weirs is a commonly used method for river enhancement in Europe (Kroes *et al.*, 2006); the lack of connectivity in terms of water, sediment, and fauna has important ecological consequences because the hydromorphological and biological conditions of the ecosystem are directly or indirectly affected (Cowx and Welcomme, 1998).

In the first step of the simulation, we analysed the effect of increasing CWB on NFSR. The simulation was implemented in the river segment between the Manchega and Torcío weirs, downstream of Alarcón dam (Fig. 2); there we simulated the removal of three small disused weirs: Carrasco, La Marmota, and Los Pontones. Weir removal generates changes in the proportion of hydromorphological units, but there was no hydraulic model available to predict such changes in this segment. Therefore, in the second step of the simulation we analysed the sensitivity of NFSR to an increase in riffle proportion (10, 20, 30, 40, and 50%) with respect to the observed values in the target river segment (Fig. 2), where there is only one native species. This segment belongs to the MCM ecotype. The Puente de Tamayo site (Cabriel River, species richness = 5) was the reference site of this fluvial ecotype (CEDEX, 2005) that we used to compare the richness in the target river segment. This is the only reference site of this ecotype in the Júcar River.

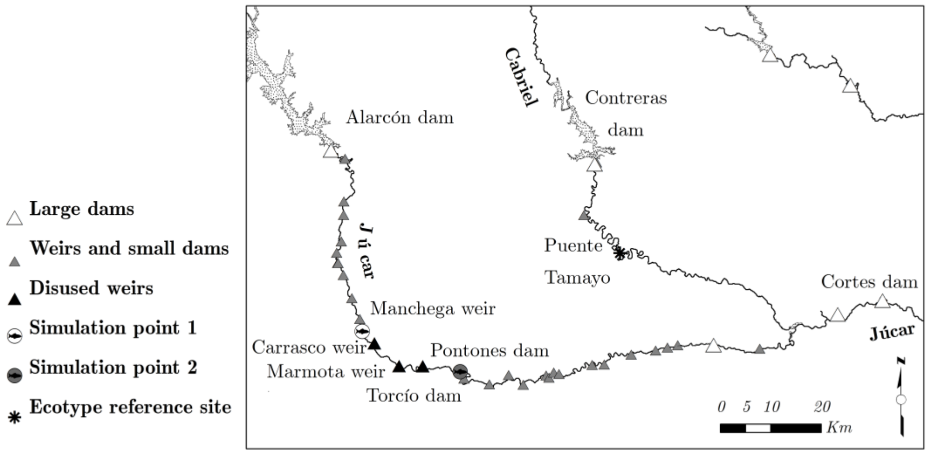

Fig. 2. Main stem of the Júcar and Cabriel Rivers, with the weirs and small dams (triangles), the 2 sampling points in the segment of simulation where the effects of mitigation measures were simulated (fish symbol). The reference site of this ecotype is also indicated (Cabriel River).

2.3 RESULTS

The variables DHS, WAT, ALT and DRA have a strong correlation (Fig. 3). Following the literature, DRA has the highest ecological interpretability for fish richness (Filipe *et al.,* 2010; Matthews and Robison, 1998; Oberdorff *et al.,* 1995); consequently we remove the first three variables as potential predictive variables. Since CON and FCV are highly correlated, we excluded the variable CON as a potential predictive variable. According to Figure 3, FIA and FMA are highly correlated, but we preserve both of them because they are important for Mediterranean fish life (Granado-Lorencio, 2000; Hermoso and Clavero, 2011).

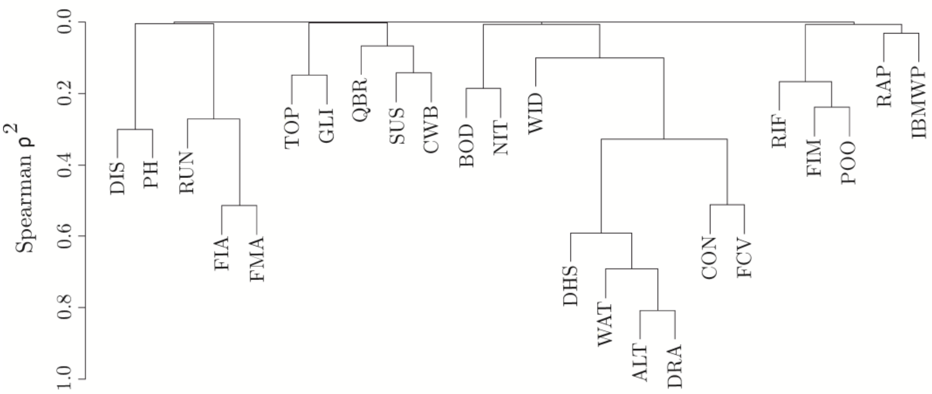

Fig. 3. Representation of the hierarchical clustering using squared Spearman correlation (ρ2) on the environmental variables, in order to indicate their similarities.

The cross-correlation analysis and the forward stepwise method allowed us to determine the key variables for predicting the NFSR. These were the QBR, RIF, CWB, drainage area (DRA), coefficient of variation of mean monthly flows (FIM), mean annual flow rate (FMA), and the IBMWP index. With these variables the ANN model was built and developed.

The architecture of the best ANN model (i.e. best performance) for NFSR consisted of three layers (7-6-1): one input layer with seven nodes (seven environmental variables); one hidden layer with six nodes; and the output layer with one output node (Fig. 4). The use of the hyperbolic tangent as the transfer function in the hidden layer and linear transformation in the output layer improved the ANN performance compared with the model using the logistic function in both layers. Therefore, both the hyperbolic tangent and linear transformation functions were used to construct the final ANN. Among the different trials of *k*-fold cross-validations, only those with k values smaller than 6 improved the network performance. Consequently, the *k*-6 was used to construct and validate the ANN.

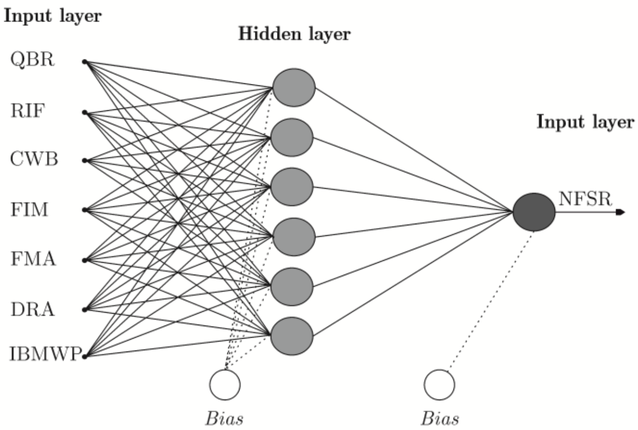

Fig. 4. Structure of the three-layered feed-forward artificial neural network with the best performance. Seven input nodes correspond to the independent environmental variables (QBR=riparian forest quality index, RIF=riffle proportion, CWB=channel length without artificial barriers, FIM = coefficient of variation of mean monthly flow, FMA = mean annual flow rate, DRA = drainage area and IBMWP index), six nodes constitutes the hidden layer and one output node shows the estimate of native fish richness.

The *r* value of the ANN model in the training procedure was 0.90 ($P < 0.05$), and it was 0.81 in the validation procedure ($P < 0.05$). The MSE values for training and validation were 0.35 and 0.62, respectively. The *r* and the MSE showed a good fit of the values estimated by the model to the observed data. Figure 5 shows the relationship between the observed values and those estimated by the model. The partial derivatives method established that the variables that contributed most to the model were IBMWP, RIF, and FMA, with a relative importance of 20.72%, 20.18%, and 16.44%, respectively (Fig. 6). This is consistent with reality, because observed species richness tend to increase for high values of IBMWP, RIF and

FMA in each of the three rivers, and the observed richness decrease for low values of these variables.

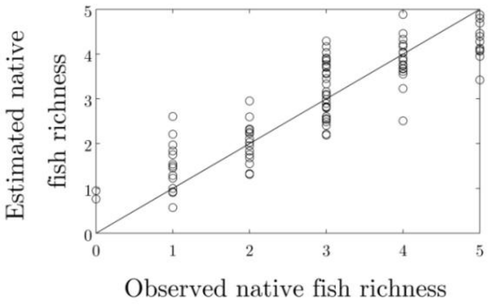

Fig. 5. Relation between the observed and estimated values of native fish species richness (NFSR) for the dataset used in the training and validation of the ANN model of the best performance (N = 90). The mean squared errors for training and validation were 0.35 and 0.62, respectively.

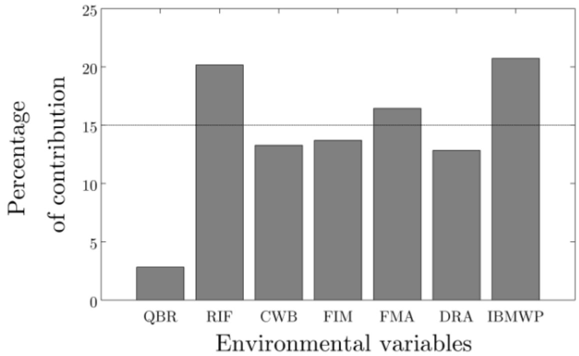

Fig. 6. Relative contribution of the environmental variables to the modelling of native fish richness, estimated by the partial derivative algorithm (PaD). The line represents the minimal significance level (15%) accordingly to Brosse (2003). See codes of variables in Figure 4.

Figure 7 illustrates the influence of the key environmental variables in predicting the richness of native fish in the model, as calculated using PaD sensitivity analysis. The positive values on the Y axis indicate a positive relationship between the input and output variable (NFSR), and negative values represent a negative influence. The values of the partial derivatives of the NFSR relative to IBMWP, RIF, and FMA are mainly positive, as indicated by the lateral histograms in the figure. This means that an increase in one of the three variables leads to an increase in the NFSR in the study area.

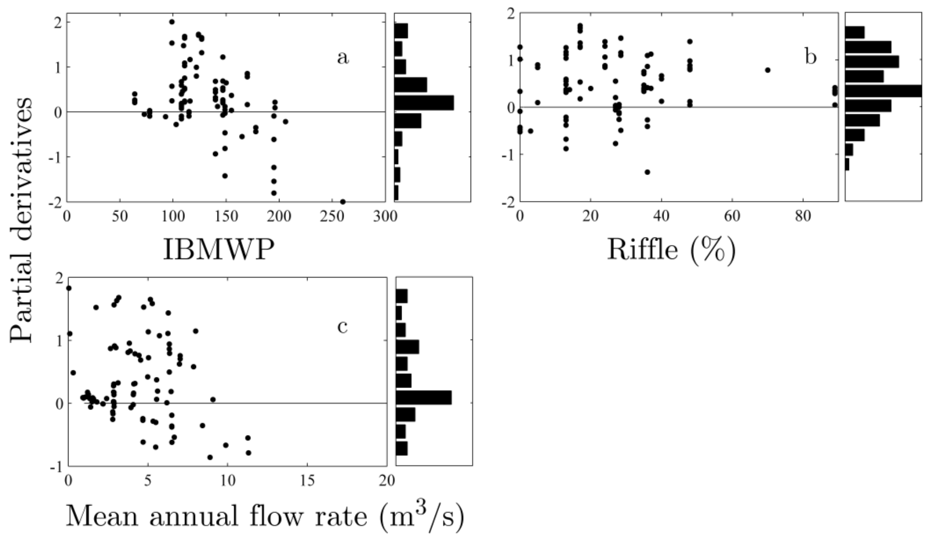

Fig. 7. Partial derivatives of the richness of native fish (NFSR) regarding IBMWP index, riffle proportion (RIF) and mean annual flow rate (FMA).

The first step of the simulation indicated that elimination of the three weirs would result in a 37 km long reach free of barriers. The model predicted that an increase of NFSR from 1 to 3 species would occur in the two simulation sites. The increase in RIF (10, 20, 30, 40, and 50%) over the observed values at each site (35% and 27%) produced a progressive increase in the richness of native fish (Fig. 8). In the first simulation site it meant an increase of richness to a constant value of 5 species. In the second simulation site the richness increased progressively until it reached 4 native species. These values are very near or match the maximum NFSR evaluated in the reference site of this river ecotype (CMM).

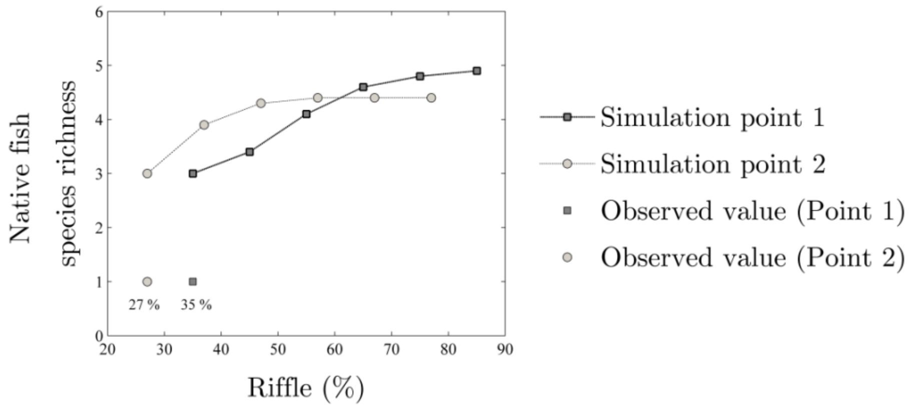

Fig. 8. Progression of the native fish species richness simulated with regular increments in the riffle proportion (RIF) over the original values estimated in the field (35 and 27 % in sites 1 and 2, respectively).

2.4 DISCUSSION

The ANN developed in this study, which predicted the NFSR based on seven environmental variables, performed well. The accuracy of the model is related to the ability of the ANN to represent the structure and non-linear processes found in nature (Drew *et al.,* 2011). The combination of techniques used during model development, such as the forward stepwise method to incorporate predictive variables (Gevrey *et al.,* 2003), the testing of different numbers of neurons (hidden layer), the comparison between transfer functions, and the testing of different *k* values (from 3 to 10) in the *k*-fold cross-validation, supported the reliability and robustness of this ANN model. The final group of key variables, with ecological relevance and consequences for river restoration, was an important result of this research, as it is discussed below.

The most significant variables identified in this study (IBMWP, RIF, and FMA) are considered to be very important for the development of Mediterranean fish communities (Bernardo *et al.,* 2003; Granado-Lorencio, 1996, 2000; Oliva-Paterna *et al.,* 2003). Among these variables, water quality usually affects aquatic species at a larger scale compared to the other variables, and large-scale studies frequently find significant associations between fish communities and abiotic factors (Jackson *et al.,* 2001). Many researchers have noted the importance of water chemistry, especially dissolved oxygen and pH, for aquatic communities (Jackson *et al.,* 2001). However, none of these variables was selected for inclusion in the final model, and instead IBMWP was used as an integrative indicator of water quality. In Spain, this index has been widely used for evaluating water quality and for ecological monitoring (Carballo *et al.,* 2009). We found that in the rivers being studied, the NFSR increases with the IBMWP (Fig. 7a); this result is consistent with results of other studies because the index is positively related to different aspects of the ecological status of rivers. For instance, in Iberian rivers Pardo *et al.* (2002) found a significant positive correlation between the IBMWP and the index of habitat diversity, and Benejam *et al.* (2010) reported that water pollution decreases the condition and fecundity of freshwater fish. Abiotic factors are also important; for example, water temperature can limit the range of species over a broad geographic scale (Shuter *et al.,* 1980), and it is a relevant factor at finer scales as well (e.g., Cunjak and Linnansaari, 2011; Grossman and Freeman, 1987). This factor was integrated into the analysis as mean annual temperature, but it ultimately was not selected for inclusion in the model. This was the only robust variable that could be calculated from the available public data. However, the pattern of seasonal variations in temperature and the changes at the scale of hydromorphological units are affected by the proportions of habitats (e.g., riffles and pools), the presence of weirs or other obstacles, and tributaries, and we believe that this factor requires further research at smaller spatial and temporal scales.

The ANN model developed at the basin scale in this study is the first step needed to develop more complex simulations at different spatial and temporal scales and to assess the effects of water scarcity and global change on Mediterranean fish communities. In the Mediterranean, water scarcity is one of the consequences of global change (García-Ruiz *et al.,* 2011), and its potential effects on aquatic systems are an issue of major concern; for instance, the reduction of water quality may pose severe risks to ecosystem integrity (Petrovic *et al.,* 2011). One future step may be to develop another ANN to include as inputs the water quality variables available from the monitoring networks in order to investigate specifically the relationships between water quality and fish richness. To date, the absence of these variables in the final ANN model suggests a low sensitivity of richness to these variables individually or in combination in comparison with other environmental factors.

Although the currently available analytical methodologies can detect the majority of relevant pollutants at their environmental levels, their dynamic environmental fate and their negative effects on the ecosystem are still poorly understood (Petrovic *et al.*, 2011). Therefore, understanding the response of the biota to both water scarcity and poor water quality is a challenging topic of research (Barceló and Sabater, 2010). In our opinion, the integration of “standard” parameters of water quality and emerging contaminants into fish distribution models is a feasible approach that might improve our understanding of the ecological responses and the effects of multiple stressors. In the regional context, the future development of richness models also should include analysis of ecological interactions at a time scale smaller than a month and incorporate parameters relevant to extreme hydrological events, such as floods and droughts. For instance, the low flows associated with water scarcity affect biogeochemical processes, decrease the dilution capacity of nutrient loads, and also decrease the natural ability of river biota to process sewage waters (Petrovic *et al.*, 2011). Spain is one of the first countries in terms of river regulation and large dams (WCD, 2000), and downstream of the reservoirs the effects of drought on community composition and structure can be intensified as a result of the competition between human uses of water and environmental values (Boix *et al.*, 2010).

In our model, the proportion of riffle habitat was the most important variable together with the IBMWP; the sensitivity analysis indicated that a positive relationship exists between RIF and the NFSR (Fig. 7b). These results are in accordance with previous studies conducted in the Iberian Peninsula, which demonstrated that native cyprinid species prefer shallow riffles, slow runs, and deep pools, with higher probability-of-use in riffles, compared to other habitats, whereas the exotic species prefer the pools (Bernardo *et al.*, 2003). The relevance of incorporating mesoscale analyses to interpret fish habitat use has been demonstrated in Mediterranean brown trout (Alcaraz-Hernández, 2011; Mouton *et al.,* 2011) and Júcar nase (Costa *et al.,* 2012) populations in Mediterranean rivers. The riffles are very important for fish reproduction; in the Júcar and Turia Rivers, gravel is the dominant substrate in riffles, and it is necessary for the spawning of lithophilic fish species such as the Iberian barbel and the Iberian chub (Doadrio, 2001b). The importance of gravel bars also has been demonstrated for other European species (e.g. barbel and dace) (Copp *et al.*, 1991). Another example is the European chub, which is considered to be a lithophilic species; it selects spawning sites with a water depth between 0.1 and 0.3 m, a stream velocity of 0.15–0.7 m s^{-1}, and a gravely substratum with grain size > 5 mm (Fredrich *et al.,* 2003).

Regarding the physical characteristics of riffles, in this study we defined them as shallow water with ripples on the surface, an average water velocity < 0.4 $m \cdot s^{-1}$, nearly symmetrical cross-sections, and a mean depth similar in magnitude to the mean substrate size (Alcaraz-Hernández, 2011). Riffles are also important for recruitment and serve as important nursery habitats (Baras *et al.,* 1996). The young of the year of some cyprinid fish species select microhabitats characterised by shallow water with low velocity (Copp, 1997; Lamouroux *et al.*, 1999; Martínez-Capel and Garcia de Jalón, 1999), which partially corresponds to riffles. The small fish prefer riffles because they provide suitable shelter from predators as well as food availability (Ilhéu *et al.,* 1999). Concerning salmonid species, Alcaraz-Hernández (2011) found a significant correlation between the density of small trout (age 0+ and 1+) and the proportion of medium substrate, which is the most abundant substrate in the riffles of four Mediterranean trout rivers.

One limitation in our study was the lack of a hydraulic model calibrated in the target river segment, which might provide estimations of habitat change after the dam removal, in terms of depth and velocity (e.g. Fjeldstad *et al.*, 2012); in spite of the short-term improvement in the hydraulic conditions, the old gravel and cobbles are now under a layer of silt, and the recovery of the area for spawning would require longer periods and suitable river flows. The hydraulic simulation could be carried out with a 1-dimentional standard model, but the application of 2D/3D models is well suited for analyses of changes after habitat rehabilitation (Alfredsen *et al.*, 2004), as it was proved in previous studies (García de Jalón and Gortázar, 2007). In the case of these old weirs in the Júcar River, where the habitats are very affected by sedimentation, the composition of the sediments can be very relevant, as well as the changes in water quality after rehabilitation; in the Júcar river there is a water quality model calibrated at the large scale, GESCAL (Paredes-Arquiola *et al.*, 2010), however its application at small scale in this case requires further technical work.

The results of PaD analysis highlighted the importance of the mean annual flow, a key variable for river habitats that greatly influences the organisation of aquatic communities (Walker and Thoms, 1993; Welcomme, 1980). The magnitude of monthly and annual flows, as well as their variability and rate of change, is presumably important for the maintenance and regeneration of riverine habitats and native biological diversity (Richter *et al.*, 1997). Specifically, Mediterranean fish species have developed optimal adaptive strategies for their survival in changing environments (Granado-Lorencio, 1996); these strategies include a short life span, rapid growth rate, high fecundity, early sexual maturity and spawning, and generalist and opportunistic feeding strategies (Ferreira *et al.*, 2007; Granado-Lorencio, 2000; Vila-Gispert *et al.*, 2002). Previous studies demonstrated that mean annual flow is a critical variable in the classification of Mediterranean rivers, thus in the interpretation of the spatial patterns of the aquatic Mediterranean communities; mean annual discharge, percentage of months with zero flow and coefficient of variation in mean annual flows represent the major gradients of variation in the Mediterranean flow regimes (Belmar *et al.*, 2011).

The magnitude of mean flows and other flow regime parameters are expected to change with the climate in the Iberian Peninsula (CEDEX, 2011), and this in turn will affect the distribution patterns of the biological communities. For river management, it is very important to anticipate future trajectories of change and identify alternative future trajectories as a basis for adaptive management and river restoration (Gregory, 2008). To evaluate whether changes would produce drawbacks or retrogression in fish diversity after restoration actions, we estimated the potential effect of a reduction in the FMA in the target segment, assuming that other relevant variables would remain the same. The last report on climate change in Spain estimated a 10–25% reduction of FMA in the Júcar River Basin (CEDEX, 2011). Based on the optimal situation after improvements in connectivity and riffle proportion, the FMA input was reduced by 10, 20, and 25%. The outputs of the model indicated that the NFSR would be reduced from 4 to 3 and from 5 to 3 in the study sites. Although this is a coarse estimation of the potential effect of reduced FMA, it is indicative of the importance of a proactive attitude towards river restoration, especially considering that efforts to improve fish communities can be overwhelmed by global change if there are no other measures of compensation. In general, in the future we can expect lower available discharges from dams to meet water demands, thus it is very important to adapt the actual water management strategies to address correctly the consequences of global change (García-Ruiz *et al.,* 2011).

In Mediterranean rivers, the hydrological alteration that occurs below dams is usually very relevant in magnitude; in many rivers, regulated flow regimes now present maximums in summer and minimums in winter, with droughts becoming more frequent and long lasting (Belmar *et al.*, 2010; Vidal-Abarca *et al.*, 2002). Management decisions must be made to handle the intense competition between humans and fish for the fresh water supply (Moyle, 1995), keeping in mind that flow diversion is one of the most important factors affecting the potential extinction of fish in Mediterranean rivers (Smith and Darwall, 2006). Because of the intense flow regulation in the Júcar, Cabriel, and Turia Rivers, it is necessary to implement suitable environmental flows that imitate the natural pattern and variability of the natural flow regime, with minimum flows in summer and hydrological events during the rainfall seasons (Arthington *et al.*, 2006), in order to promote the integrity and sustainability of the freshwater ecosystems. Existing water diversions have led to the replacement of lotic habitats by lentic habitats, which in turn has caused loss of the majority of distinctive fish associations in Mediterranean rivers; these associations have been replaced by other communities adapted to the new ecological conditions (Granado-Lorencio, 1996). Thus, implementation of environmental flow regimes in river restoration is an essential mitigation measure for improving the ecological status of the rivers (*sensu* the WFD).

Although mean annual flow was analysed in this study, it is necessary to emphasise that environmental flow regimes cannot be defined solely in terms of mean monthly or mean annual flows because flow variability is crucial for the maintenance of native communities. Overall, maintenance of natural flow variability is an important principle for riverine ecosystem protection and restoration (Jowett and Biggs, 2009; Poff *et al.*, 2010). In this study, one of the limitations in the data was the monthly time scale of the flow rate and the derived variables; other indicators of the hydrological regime (e.g., the ramping rate during peak flows and duration of floods and droughts) can provide relevant ecological information, but they could not be robustly estimated using monthly flow data. Therefore, the ANN model developed herein is the first step in approaching more complex simulations at smaller time scales (e.g., based on daily flow data). Although daily flow data are not yet available for the 90 sampling sites, a new model could be developed with a smaller database consisting of daily flow data and indicators of hydrological alteration (Richter *et al.*, 1997). The ANN at the basin scale then could be compared or validated over smaller temporal and spatial scales, and the effects of water scarcity could be estimated with higher reliability. Despite the inconvenience of having only monthly data, this ANN model may be useful for environmental flow assessments because it is a model of ecological response to habitat and hydrological changes. In general, this model can be viewed as a hydromorphological model for predicting NFSR at the basin scale because it relates hydromorphological variables to the fish communities. Therefore, this model may be integrated in methodological frameworks of environmental flows, e.g., in the IFIM (Instream Flow Incremental methodology, Bovee *et al.*, 1988), Eloha (Poff *et al.*, 2010), and other recent approaches at the basin scale (Paredes-Arquiola *et al.*, 2011 in press).

The simulation in the Júcar River indicated that richness increases with CWB and RIF (Fig. 8), which agrees with results of the PaD sensitivity analysis (Fig. 7b). These results also agree with those from previous studies in the Iberian Peninsula. For example, Alexandre and Almeida (2010) observed that fish richness was higher at the sites where direct influence of artificial transverse barriers was smaller. Barriers, weirs, or dams disrupt the longitudinal continuity of the river flow and sediments and make the migratory movements of fish difficult or impossible (Meixler *et al.,* 2009). The construction of dams and flow regulation have also created favourable conditions for invasive fish species (Corbacho and Sánchez, 2001; Poulet, 2007; Vila-Gispert *et al.,* 2005) and are considered to be a major factor in the dramatic reduction of native rheophilic species, which depend on riffles for reproduction (García de Jalón *et al.,* 2007). Poor river connectivity has been identified as one of the main causes of declines in many continental Iberian fish species (Casals, 2005; Lucas and Baras, 2001; Santo, 2005) and European species (Kroes *et al.,* 2006; Marmulla, 2001). Juvenile cyprinids may be especially vulnerable to barriers and other local alterations; thus, maintaining connectivity and local habitat quality are extremely important for supporting native fish populations (Santos *et al.,* 2011). Accordingly, the WFD requires effective passage and undisturbed migration of fish as a key component to restore and manage watersheds (European Commission, 2000). In this study, the proposal for removing weirs is based on the current knowledge of the river reaches, which contain abandoned and obsolete structures whose water rights are not in use. Weir removal and the legal process of water rights cessation are important tools for river restoration at the basin scale, and they should be widely applied to improve the status of Mediterranean aquatic communities.

In the segment of simulation the potential fish community after the mitigation measures could be assessed, as follows. The only native species actually in that segment is *L. guiraonis*. In the Júcar River between the two large reservoirs of Alarcón and Cortes (see Fig. 1), other native species (*I. alburnoides*, *C. paludica*, *S. pyrenaicus* and *A. anguilla*) can be found; thus, these would be the most probable species to colonise the segment. The reference site of the ecotype (MCM) is in the Cabriel River, with a potential connectivity to the target segment through a large reservoir; however, obstacles are present that block this connectivity and make it impassable for fish, especially in the Júcar River. Other species that potentially would colonise the target segment from the Cabriel River are *A. anguilla, S. fluviatilis, and P. arrigonis.* Therefore, the removal of weirs could improve the situation for two critically endangered fish species and contribute to meeting the European Recovery Plan for the Eel (Regulation 1100/2007; European

Commission, 2007). River longitudinal connectivity is also extremely important for maintaining the conservation status of many freshwater species included in the Nature 2000 Network (Habitats Directive 92/43/EEC; European Commission, 1992), as was highlighted in previous studies (Ordeix *et al.*, 2011). Considering the maximum value of NFSR = 5 obtained by modelling, this is the maximum potential value according to the reference site, in the case that the proportion of riffles could reach the percentages specified (Fig. 8). However, the gradient of the river channel could limit the final result of the mitigation measures in some river reaches in terms of the RIF. In the reference sites of the Júcar River Basin, 70% was the highest recorded riffle percentage; thus, according to the simulation, the maximum richness we could expect would be between 4 and 5 in the best situation.

One of the limitations of the ANN model is that this maximum of 5 fish species was estimated based in the input variables of the model without consideration of the environmental conditions (i.e., upstream and downstream) around the target segment. Therefore, due to the lack of river connectivity with other segments with higher diversity, more actions would be necessary to reach such a value of 5; these actions may include the removal of other barriers and/or the allocation of fish by the responsible administration. Another limitation of the study is that we did not consider the biological interactions (e.g., food availability and inter-species competition). The target segment contained four exotic species of fish (*Gobio gobio*, *Alburnus alburnus*, *Lepomis gibbosus*, and *Micropterus salmoides*), which could interfere with the recovery of the NFSR because they can compete with or predate on the native fish. The importance of habitat and exotic species for the recovery of native populations is undoubtedly an important issue that must be considered in these Mediterranean rivers.

The final ANN model combines variables describing physical habitat and water quality, and it contributes to identifying the primary drivers of the NFSR patterns in Mediterranean rivers; additionally, consideration of variables associated with different spatial scales provides a high potential for model transferability (Leftwich *et al.*, 1997) to other rivers and basins. Once the importance of longitudinal connectivity and riffle proportion for the NFSR has been demonstrated, these variables should be considered in river restoration strategies and projects.

Therefore, the model and the results described herein may support technical decisions for the management and ecological restoration of Mediterranean rivers.

Acknowledgements

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the projects SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and POTECOL "Evaluación del Potencial Ecológico de Ríos Regulados por Embalses y Desarrollo de Criterios para su mejora según la Directiva Marco del Agua" (CGL2007-66412). We thank to Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment) for the data provided to develop this study. We thank to Sasa Plestenjak in the collaboration for building the first fish database elaborated in this research.

References

Aguilella A, Riera J, Gómez-Serrano MA, Mayoral O, Moreyra E. Evaluación del estado ecológico de los ríos de la cuenca hidrográfica del Júcar mediante el uso del índice QBR. Valencia: Jardí Botànic, Universitat de València; 2005.

Alba-Tercedor J, Sánchez-Ortega A. Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de Hellawell (1978). Limnetica 1988; 4: 51-56.

Alcaraz-Hernández JD. Estado de las poblaciones de trucha común en ríos de la Comunidad Valenciana y caracterización de sus hábitats. Valencia: Universitat Politecnica de Valencia; 2011.

Alexandre CM, Almeida PR. The impact of small physical obstacles on the structure of freshwater fish assemblages. River Res. Appl. 2010; 26: 977-994.

Alfredsen K, Borsanyi P, Harby A, Fjeldstad H-P, Wersland S-E. Application of habitat modelling in river rehabilitation and artificial habitat design. Hydroécol. Appl. 2004; 14: 105-117.

Angermeier PL, Davideanu G. Using fish communities to assess streams in Romania: Initial development of an Index of biotic integrity. Hydrobiologia 2004; 511: 65-78.

Arthington AH, Bunn SE, Poff NL, Naiman RJ. The challenge of providing environmental flow rules to sustain river ecosystems. Ecol. Appl. 2006; 16: 1311-1318.

Baras E, Philippart JC, Nindaba J. Importance of gravel bars as spawning grounds and nurseries for european running water cyprinids. In Ecohydraulics 2000: 2nd International Symposium on Habitat Hydraulics, vol. A. A. Leclerc, M, Capra, H, Valentin, S, Boudreault, A, Côté, Y, editors. INRS-Eau: Quebec; 1996. p. 367–378.

Barceló D, Sabater S. Water quality and assessment under scarcity: Prospects and challenges in Mediterranean watersheds. J. Hydrol. 2010; 383: 1-4.

Belmar O, Velasco J, Martinez-Capel F. Hydrological Classification of Natural Flow Regimes to Support Environmental Flow Assessments in Intensively Regulated Mediterranean Rivers, Segura River Basin (Spain). Environ. Manage. 2011; 47: 992-1004.

Belmar O, Velasco J, Martínez-Capel F, Marín AA. Natural flow regime, degree of alteration and environmental flows in the Mula stream (Segura River basin, SE Spain). Limnetica 2010; 29: 353-368.

Benejam L, Angermeier PL, Munné A, García-Berthou E. Assessing effects of water abstraction on fish assemblages in Mediterranean streams. Freshw. Biol. 2010; 55: 628-642.

Bernardo JM, Ilhéu M, Matono P, Costa AM. Interannual variation of fish assemblage structure in a Mediterranean river: implications of streamflow on the dominance of native or exotic species. River Res. Appl. 2003; 19: 521-532.

Bishop CM. Neural Networks for Pattern Recognition. Oxford: University Press; 1996.

Boix D, García-Berthou E, Gascón S, Benejam L, Tornés E, Sala J, et al. Response of community structure to sustained drought in Mediterranean rivers. J. Hydrol. 2010; 383: 135-146.

Bovee KD, Lamb BL, Bartholow JM, Stalnaker CB, Taylor J, Henriksen J. Stream habitat analysis using the instream flow incremental methodology: U.S. Geological Survey, Biological Resources Division Information and Technology Report USGS/BRD-1998-0004; 1988.

Brosse S, Arbuckle CJ, Townsend CR. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. Biodivers. Conserv. 2003; 12: 2057-2075.

Brosse S, Guegan J-F, Tourenq J-N, Lek S. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 1999; 120: 299-311.

Caissie D. River discharge and channel width relationships for New Brunswick rivers. Canadian Technical Report of Fisheries and Aquatic Sciences 2637, 2006, pp. 26.

Caissie D, El-Jabi N. Comparison and regionalization of hydrologically based instream flow techniques in Atlantic Canada. Can. J. Civ. Eng. 1995; 22: 235-246.

Carballo R, Cancela J, Iglesias G, Marín A, Neira X, Cuesta T. WFD Indicators and Definition of the Ecological Status of Rivers. Water Resour. Manag. 2009; 23: 2231-2247.

Casals F. Les comunitats íctiques dels rius mediterranis: relació amb les condicions ambientals. Barcelona: Universidad de Barcelona; 2005.

CEDEX. Caracterización de los tipos de ríos y lagos. Versión 4. Madrid: Centro de Estudios Hidrográficos del CEDEX; 2005.

CEDEX. Evaluación del impacto del cambio climático en los recursos hídricos en régimen natural. Resumen ejecutivo. Available at: http://marm.es/es/agua/formacion/. Madrid: Ministerio de Medio Ambiente y Medio Rural y Marino; CEDEX; 2011.

Clavero M, Blanco-Garrido F, Prenda J. Fish fauna in Iberian Mediterranean river basins: biodiversity, introduced species and damming impacts. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2004; 14: 575-585.

Copp GH. Microhabitat use of fish larvae and 0+ juveniles in a highly regulated section of the River Great Ouse. Regul. River. 1997; 13: 267-276.

Copp GH, Oliver JM, Peňáz M, Roux AL. Juvenile fishes as functional describers of fluvial ecosystem dynamics: Applications on the river rhǒne, France. Regul. River. 1991; 6: 135-145.

Corbacho C, Sánchez JM. Patterns of species richness and introduced species in native freshwater fish faunas of a Mediterranean-type basin: the Guadiana River (southwest Iberian Peninsula). Regul. River. 2001; 17: 699-707.

Costa RMS, Martínez-Capel F, Muñoz-Mas R, Alcaraz-Hernández JD, Garófano-Gómez V. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (river Cabriel, Spain). River Res. Appl. 2011; 28: 740-752.

Cowx IG, Welcomme RL. Rehabilitation of Rivers for Fish. Oxford: Wiley; 1998.

Cunjak R, Linnansaari T. Biological significance of thermal refugia for juvenile atlantic salmon during extreme heat events in rivers. NoWPaS Conference. Gotein-Libarrenx, France; 2011.

Cuttelod A, García N, Abdul Malak D, Temple H, Katariya V. The Mediterranean: a biodiversity hotspot under threat. In: Vié J-C, Hilton-Taylor C, Stuart SN, editors. The 2008 Review of The IUCN Red List of Threatened Species. IUCN Gland, Switzerland, 2008.

Cheng L, Lek S, Lek-Ang S, Li Z. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 2012; 42: 127-136.

CHJ. La ictiofauna como elemento de calidad de los ríos de la demarcación hidrográfica del río Júcar. Valencia: Confederación Hidrográfica del Júcar; 2007.

Dedecker A, Goethals P, D'Heygere T, Gevrey M, Lek S, De Pauw N. Application of artificial neural network models to analyse the relationships between Gammarus pulex L. (Crustacea, Amphipoda) and river characteristics. Environ. Monit. Assess. 2005; 111: 223-241.

Dimopoulos I, Chronopoulos J, Chronopoulou-Sereli A, Lek S. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecol. Model. 1999; 120: 157-165.

Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 1995; 2: 1-4.

Doadrio I. Atlas y Libro Rojo de los Peces Continentales de España. Madrid: Museo Nacional de Ciencias Naturales; 2001.

Dolloff CA, Hankin DG, Reeves GH. Basinwide estimation of habitat and fish populations in streams: U.S. Department of Agriculture. Forest Service. Southeastern Forest Experiment Station; 1993.

Dormann, C.F., 2011. Modelling species' distributions. In: Jopp, F., Reuter, H. and Breckling, B. (eds.), Modelling complex ecological dynamics: an Introduction into ecological modelling for students, teachers and scientists, Springer-Verlag, Berlin, 179-196.

Drew CA, Wiersma Y, Huettmann F. Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. New York: Springer; 2011.

Elith J, H. Graham C, P. Anderson R, Dudík M, Ferrier S, Guisan A, *et al.* Novel methods improve prediction of species' distributions from occurrence data. Ecography 2006; 29: 129-151.

Estrela T, Fidalgo A, Fullana J, Maestu J, Pérez MA, Pujante AM. Júcar Pilot River Basin, provisional article 5 report Pursuant to the Water Framework Directive. Valencia: Confederación Hidrográfica del Júcar; 2004.

European Commission. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities L 206, 22.7.1992:7; 1992.

European Commission. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy: Official Journal of the European Communities L 327: 1-72.; 2000.

European Commission. Council Regulation (EC) No 1100/2007 of 18 September 2007 establishing measures for the recovery of the stock of European eel Official Journal of the European Union 22.9.2007 L 248: 17-23 2007.

Ferreira T, Oliveira J, Caiola N, De Sostoa A, Casals F, Cortes R, *et al*. Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Manag. Ecol. 2007; 14: 473-481.

Filipe AF, Filomena Magalhães M, Collares-Pereira MJ. Native and introduced fish species richness in Mediterranean streams: the role of multiple landscape influences. Divers. Distrib. 2010; 16: 773-785.

Fjeldstad HP, Barlaup BT, Stickler M, Gabrielsen SE, Alfredsen K. Removal of weirs and the influence on physical habitat for salmonids in a Norwegian river. River Res. Appl. 2012; 28: 753-763.

Fredrich F, Ohmann S, Curio B, Kirschbaum F. Spawning migrations of the chub in the River Spree, Germany. J. Fish Biol. 2003; 63: 710-723.

Freyhof J, Brooks E. European Red List of Freshwater Fishes. Luxembourg: Publications Office of the European Union; 2011.

García-Ruiz JM, López-Moreno JI, Vicente-Serrano SM, Lasanta–Martínez T, Beguería S. Mediterranean water resources in a global change scenario. Earth-Sci. Rev. 2011; 105: 121-139.

García de Jalón D, Gortázar J. Evaluation of instream habitat enhancement options using fish habitat simulations: case-studies in the river Pas (Spain). Aquat. Ecol. 2007; 41: 461-474.

García de Jalón D, Sánchez Navarro R, Serrano J. Alteraciones de los regímenes de caudales de los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Gasith A, Resh VH. Streams in mediterranean climate regions: Abiotic Influences and Biotic Responses to Predictable Seasonal Events. Annu. Rev. Ecol. Syst. 1999; 30: 51-81.

Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 2003; 160: 249-264.

Goethals P, Dedecker A, Gabriels W, Lek S, De Pauw N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 2007; 41: 491-508.

Granado-Lorencio C. Ecología de peces. Sevilla: Universidad de Sevilla; 1996.

Granado-Lorencio C. Ecología de comunidades: el paradigma de los peces de agua dulce. Sevilla: Universidad de Sevilla; 2000.

Gregory S. River restoration: restoring dynamic riverine processes in a changing world or erecting monuments to our good intentions. In: Gumiero B, Rinaldi M, Fokkens B, editors. IVth ECRR International Conference on River Restoration, Servolo Island, Venice, 2008, pp. 769-778.

Grossman GD, Freeman MC. Microhabitat use in a stream fish assemblage. J. Zool. 1987; 212: 151-176.

Gutiérrez-Estrada JC, Bilton DT. A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters. Ecol. Model. 2010; 221: 1451-1462.

Hassoun M. Fundamentals of Artificial Neural Networks. Cambridge: The MIT Press; 1995.

Hattermann FF, Krysanova V, Habeck A, Bronstert A. Integrating wetlands and riparian zones in river basin modelling. Ecol. Model. 2006; 199: 379-392.

He Y, Wang J, Lek-Ang S, Lek S. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. Sci. Total Environ. 2010; 408: 4211-4220.

Hermoso V, Clavero M. Threatening processes and conservation management of endemic freshwater fish in the Mediterranean basin: a review. Mar. Freshwater Res. 2011; 62: 244-254.

Ibarra AA, Gevrey M, Park Y-S, Lim P, Lek S. Modelling the factors that influence fish guilds composition using a back-propagation network: Assessment of metrics for indices of biotic integrity. Ecol. Model. 2003; 160: 281-290.

Ilhéu M, Costa AM, Bernardo JM. Habitat use by fish species in a Mediterranean temporary river: the importance of riffles. Proceedings of the 3rd International Symposium on Ecohydraulics. Salt Lake City: Utah State University; 1999.

Isa IS, Omar S, Saad Z, Osman MK. Performance comparison of different multilayer perceptron network activation functions in automated weather classification. Proceedings of the 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation. Kota Kinabalu, Malaysia; 2010. p. 71-75.

Jackson DA, Peres-Neto PR, Olden JD. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 2001; 58: 157–170.

Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, Teachers and Scientists. Berlin: Springer-Verlag; 2011.

Jowett IG, Biggs BJF. Application of the ‘natural flow paradigm’ in a New Zealand context. River Res. Appl. 2009; 25: 1126-1135.

Karr JR. Assessment of Biotic Integrity Using Fish Communities. Fisheries 1981; 6: 21-27.

Karul C, Soyupak S, Çilesiz AF, Akbay N, Germen E. Case studies on the use of neural networks in eutrophication modeling. Ecol. Model. 2000; 134: 145-152.

Knudby A, LeDrew E, Brenning A. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. Remote Sens. Environ. 2010; 114: 1230-1241.

Kroes MJ, Gough PP, Wanningen H, Schollema P, Ordeix M, Vesely D. From sea to source. Practical guidance for the restoration of fish migration in European Rivers. Interreg IIIC Project “Community Rivers”. Groningen, The Netherlands; 2006.

Kurková V. Kolmogorov's theorem and multilayer neural networks. Neural Netw. 1992; 5: 501-506.

Lamouroux N, Olivier J-M, Persat H, PouilLy M, Souchon Y, Statzner B. Predicting community characteristics from habitat conditions: fluvial fish and hydraulics. Freshw. Biol. 1999; 42: 275-299.

Laws EA. Aquatic Pollution: An Introductory Text. United States of America: John Wiley & Sons; 2000.

Leftwich KN, Angermeier PL, Dolloff CA. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Trans. Am. Fish. Soc. 1997; 126: 725-734

Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 1996; 90: 39-52.

Lek S, Scardi M, Verdonschot P, Descy JP, Park YS, editors. Modelling Community Structure in Freshwater Ecosystems. Berlin: Springer-Verlag; 2005.

Leopold LB, Maddock T. The hydraulic geometry of stream channels and some physiographic implications. U.S. Geological Survey Professional Paper 252. Washington: United States Government Printing Office; 1953.

Leopold LB, Wolman MG, Miller JP. Fluvial Processes in Geomorphology. San Francisco: W.H. Freeman and Company; 1964.

Lucas MC, Baras E. Migration of Freshwater Fishes. Oxford, United Kingdom: Blackwell Science; 2001.

Maier HR, Dandy GC. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environ. Modell. Softw. 2000; 15: 101-124.

Marmulla G. Dams, fish and fisheries: Opportunities, challenges and conflict resolution. FAO Fisheries Technical Paper No. 419. Rome: FAO; 2001.

Martínez-Capel F, Garcia de Jalón D. Desarrollo de Curvas de preferencia de microhábitat para Leuciscus pyrenaicus y Barbus bocagei por buceo en el río Jarama (Cuenca del Tajo). Limnetica 1999; 17: 71-83.

Mastrorillo S, Dauba F, Oberdorff T, Guégan J-F, Lek S. Predicting local fish species richness in the garonne river basin. Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 1998; 321: 423-428.

Matthews WJ, Robison HW. Influence of Drainage Connectivity, Drainage Area and Regional Species Richness on Fishes of the Interior Highlands in Arkansas. Am. Midland Nat. 1998; 139: 1-19.

Meixler MS, Bain MB, Todd Walter M. Predicting barrier passage and habitat suitability for migratory fish species. Ecol. Model. 2009; 220: 2782-2791.

Mouton A, Dedecker A, Lek S, Goethals P. Selecting Variables for Habitat Suitability of Asellus (Crustacea, Isopoda) by Applying Input Variable

Contribution Methods to Artificial Neural Network Models. Environ. Model. Assess. 2010; 15: 65-79.

Mouton AM, Alcaraz-Hernández JD, De Baets B, Goethals PLM, Martínez-Capel F. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Modell. Softw. 2011; 26: 615-622.

Moyle PB. Conservation of native freshwater fishes in the Mediterranean-type climate of California, USA: A review. Biol. Conserv. 1995; 72: 271-279.

Munné A, Prat N, Solà C, Bonada N, Rieradevall M. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2003; 13: 147-163.

Naiman RJ, Decamps H, Pollock M. The Role of Riparian Corridors in Maintaining Regional Biodiversity. Ecol. Appl. 1993; 3: 209-212.

Oberdoff T, Guégan J-F, Hugueny B. Global scale patterns of fish species richness in rivers. Ecography 1995; 18: 345-352.

Oberdorff T, Guégan J-F, Hugueny B. Global scale patterns of fish species richness in rivers. Ecography 1995; 18: 345-352.

Olden JD, Jackson DA. A comparison of statistical approaches for modelling fish species distributions. Freshw. Biol. 2002a; 47: 1976-1995.

Olden JD, Jackson DA. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 2002b; 154: 135-150.

Olden JD, Lawler JJ, Poff NL. Machine Learning Methods Without Tears: A Primer for Ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Olden JD, Poff NL, Bledsoe BP. Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. Ecol. Inform. 2006; 1: 33-42.

Oliva-Paterna FJ, Miñnano PA, Torralva M. Habitat quality affects the condition of Barbus sclateri in Mediterranean semi-arid streams. Environ. Biol. Fishes 2003; 67: 13-22.

Ordeix M, Pou-Rovira Q, Sellarès N, Munné A, Bardina M, Casamitjana A, *et al*. Fish pass assessment in the rivers of Catalonia (NE Iberian Peninsula). A case study of weirs associated with hydropower plants and gauging stations. Limnetica 2011; 30: 405-426.

Özesmi SL, Tan CO, Özesmi U. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 2006; 195: 83-93.

Pardo I, Álvarez M, Casas J, Moreno JL, Vivas S, Bonada N, *et al*. El hábitat de los ríos mediterráneos. Diseño de un índice de diversidad de hàbitat. Limnetica 2002; 21: 115-132.

Paredes-Arquiola J, Andreu-Álvarez J, Martín-Monerris M, Solera A. Water Quantity and Quality Models Applied to the Jucar River Basin, Spain. Water Resour. Manag. 2010; 24: 2759-2779.

Paredes-Arquiola J, Martinez-Capel F, Solera A, Aguilella V. Implementing environmental flows in complex water resources systems–case study: the Duero river basin, Spain. River Res. Appl. 2011: n/a-n/a.

Patten D. Riparian ecosytems of semi-arid North America: Diversity and human impacts. Wetlands 1998; 18: 498-512-512.

Petrovic M, Ginebreda A, Acuña V, Batalla RJ, Elosegi A, Guasch H, *et al.* Combined scenarios of chemical and ecological quality under water scarcity in Mediterranean rivers. Trends Anal. Chem. 2011; 30: 1269-1278.

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, *et al.* The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshw. Biol. 2010; 55: 147-170.

Poulet N. Impact of weirs on fish communities in a piedmont stream. River Res. Appl. 2007; 23: 1038-1047.

Quinn JM, Boothroyd IKG, Smith BJ. Riparian buffers mitigate effects of pine plantation logging on New Zealand streams: 2. Invertebrate communities. For. Ecol. Manage. 2004; 191: 129-146.

Recknagel F. Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation. Berlin (Germany) and New York: Springer-Verlag; 2003.

Reyjol Y, Hugueny B, Pont D, Bianco PG, Beier U, Caiola N, *et al.* Patterns in species richness and endemism of European freshwater fish. Glob. Ecol. Biogeogr. 2007; 16: 65-75.

Richter B, Baumgartner J, Wigington R, Braun D. How much water does a river need? Freshw. Biol. 1997; 37: 231-249.

Rosenberg DM, Resh VH, editors. Freshwater biomonitoring and benthic macroinvertebrates. New York: Chapman & Hall; 1993.

Santo M. Dispositivos de passagem para peixes em Portugal. Direcçao-Geral dos Recursos Florestais Lisboa: Editideias - Ediçao e Produçao, Lda.; 2005.

Santos J, Reino L, Porto M, Oliveira J, Pinheiro P, Almeida P, *et al.* Complex size-dependent habitat associations in potamodromous fish species. Aquat. Sci. 2011; 73: 233-245.

Shepherd AJ. Second-Order Methods for Neural Networks. New York: Springer-Verlag; 1997.

Shuter BJ, Maclean JA, Fry FEJ, Regier HA. Stochastic Simulation of Temperature Effects on First-Year Survival of Smallmouth Bass. Trans. Am. Fish. Soc. 1980; 109: 1-34.

Smith KG, Darwall WRT. The status and distribution of freshwater fish endemic to the Mediterranean Basin: IUCN, Gland, Switzerland and Cambridge, UK.; 2006.

Soria E, Martín JD, Martinez M, Magdalena JR, Serrano AJ, editors. Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. New York: Information Science Reference; 2010.

Tan Y, Van Cauwenberghe A. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intell. 1999; 12: 21-35.

van Jaarsveld AS, Freitag S, Chown SL, Muller C, Koch S, Hull H, *et al.* Biodiversity assessment and conservation strategies. Science 1998; 279: 2106-2108.

van Wijk MT, Bouten W. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. Ecol. Model. 1999; 120: 181-197.

Vidal-Abarca M, Suárez M, Gómez R, Moreno JL, Guerrero C. Diel variation in physical and chemical parameters in a semiarid stream in Spain (Chicamo Stream). Verh. Internat. Verein. Limnol. 2002; 28: 1111-1115.

Vidal-Abarca MR, Suárez ML, Ramírez-Díaz L. Ecology of Spanish semiarid streams. Limnetica 1992; 8: 151-160.

Vila-Gispert A, Alcaraz C, García-Berthou E. Life-history traits of invasive fish in small Mediterranean streams. Biol. Invasions 2005; 7: 107-116-116.

Vila-Gispert A, Moreno-Amich R, García-Berthou E. Gradients of life-history variation: an intercontinental comparison of fishes. Rev. Fish. Biol. Fish. 2002; 12: 417-427.

Walker KF, Thoms MC. Environmental effects of flow regulation on the lower river Murray, Australia. Regul. River. 1993; 8: 103-119.

WCD. Dams and development: A new framework for decision-making. London: Earthscan; 2000.

Welcomme RL. Cuencas Fluviales. Roma: FAO; 1980.

Wurbs R. Methods for Developing Naturalized Monthly Flows at Gaged and Ungaged Sites. J. Hydrol. Eng. 2006; 11: 55-64.

Zitek A, Schmutz S, Jungwirth M. Assessing the efficiency of connectivity measures with regard to the EU-Water Framework Directive in a Danube-tributary system. Hydrobiologia 2008; 609: 139-161.

Chapter 3

Role of invasive species and habitat degradation on freshwater native fish diversity in Mediterranean River Basins

Esther Julia Olaya-Marín, Francisco Martínez-Capel, Juan Diego Alcaraz-Hernández

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

ABSTRACT

The presence of invasive species and habitat alteration have been frequently identified as major threats to Mediterranean native freshwater fishes. However, it has been questioned the main or secondary role of invasive species as drivers of the decline of native fish species richness (NFSR); therefore, it is possible that habitat plays the main role in some aquatic ecosystems, meaning that exotic fish invasions and native fish decline could be a consequence of the habitat degradation. It is clear that the knowledge of the main drivers is very important for river restoration, in order to prioritize effective management actions. In this study, we used the multilayer feed-forward ANN to analyse the effect of habitat alteration and invasive species on native fish richness, in three Mediterranean rivers. We built three ANNs, one with biological variables, another only with habitat variables and a third model with the most relevant environmental variables of the previous models. To prevent overfitting we used together "early stopping" and cross-validation and it was used the Levenberg-Marquardt algorithm for optimization. The importance of the ANNs input variables was determined using the partial derivatives (PaD) method. We found the best performance with only habitat variables, in contrast to the model with biological variables and the third (combined) model. The third model included percentage of riffles and length of rivers without barriers as the main habitat variables, while the number of exotic predator species was the fifth in importance. The results indicate that habitat variables (including stream flow and habitat connectivity) are important drivers of NFSR. The findings of this study highlight the need to propose mitigation measures related to the improvement of habitat to conserve and restore Mediterranean rivers.

Keywords: Artificial neural networks (ANN); invasive species; fish richness; river regulation; habitat degradation; driver; passenger; Mediterranean rivers.

3.1 INTRODUCTION

In the European context, freshwater species of the Mediterranean river basins have a high level of endemism (Abell *et al.*, 2008), due to their geographical isolation and climate variability, characterized by intermittent periods of torrential rains and droughts and a high intra-annual and inter-annual variation of flow (Gasith and Resh, 1999). It is estimated that 70% of the 228 endemic freshwater Mediterranean fish species are threatened or endangered and are nearly extinct (Hermoso and Clavero, 2011; Ribeiro and Leunda, 2012; Smith and Darwall, 2006). This situation is represented in the Iberian Peninsula, where approximately 80% of the fish species are endemic (Doadrio, 2002). Therefore, the design of conservation and restoration actions is crucial to diminish biodiversity loss rates.

But, the development of new knowledge about the factors driving diversity loss is necessary to prioritize effective management actions (Hermoso *et al.*, 2011).

Some factors such as habitat alteration and invasive species have been cited by many researchers as the main threats to native freshwater fish (Didham *et al.*, 2007; García-Berthou *et al.,* 2005; Moyle, 1995). In fact, the main Iberian Peninsula rivers have more exotic than native fish species (Clavero and García-Berthou, 2006). Inside the principal ecological impacts and threats exerted by invasive species on native ones stands hybridization and genetic introgression, disease vectors introduction and parasites, competition over resources, predation and ecosystem disruption (Ribeiro and Leunda, 2012). However, the positive correlation between invasive species and the decline of native species has not been proven (Gurevitch and Padilla, 2004), therefore, the invasive species impact on the environment of native species has been questioned. Some authors have suggested that invasive species may operate as "*passengers*" and not "*drivers*" in the ecological change (Didham *et al.,* 2005; MacDougall and Turkington, 2005; Spieles, 2010). *Passenger* means that invasive species operate as a symptom of habitat modification and degradation, *driver* means that invasive species originally generate the ecological change (Spieles, 2010).

There are few studies that analyse the habitat degradation role and invasive species in the extinction of native fish species. Corbacho and Sánchez (2001) analysed the factors which affect native species richness over 30 rivers of the Guadiana River basin (southwestern of Iberian Peninsula), their study was based on biotic and hydrologic variables. Using principal component and multiple regression analysis, they found that habitat degradation could be the main cause in the decline of native species and the spread of invasive species in the Guadiana basin. Light and Marchetti (2007) compiled fish presence data, conservation status, land use and hydrologic modifications at catchment level, and designed models (using regression analysis) to examine the decline of native species in California. Their results showed that the invasions are the main *driver* of population declines and extinctions of fish. Hermoso *et al.* (2011) discussed the importance of habitat degradation and invasive species in the decline of native fish assemblages in the Guadiana River basin, using structural equations modelling (SEM). They found that the abundance of invasive species is the best predictor of native species decline, and habitat degradation does not play an active role. The factors which affect the native fish richness species are important for management and conservation of Mediterranean aquatic ecosystems, because if an invasive species is a change driver, their removal should lead to recovery of native species in either richness or abundance. By contrast, the removal effects of an invasive species would be minimal if not exert significant control in the ecosystem change (Bulleri *et al.,* 2010).

Artificial neural networks (ANNs) have been used in ecology by a large number of authors due to its adaptability to all problem type and the good results obtained in different studies (Albañez-Lucero and Arreguín-Sánchez, 2009; Brosse *et al.,* 1999; Garzón *et al.,* 2006; Hilbert and Ostendorf, 2001; Lippitt *et al.,* 2008). ANNs are mathematical models inspired by the structure and functioning of biological neural systems (Gutiérrez-Estrada and Bilton, 2010; Olden *et al.,* 2008) and have been successfully applied in freshwater fish ecological studies (Ibarra *et al.,* 2003; Mastrorillo *et al.,* 1998; Penczak, 2011; Tirelli and Pessani, 2009). This paper aims to determine the relative importance of the two main factors in the reduction of native fish species richness (NFSR), i.e. habitat and ecological interactions among aquatic species, in the Júcar, Cabriel and the Turia rivers. For this purpose, we created three different ANN models of native fish richness based on three sets of variables: biological, habitat and variables describing the natural environmental variability in the river basin. The predictive variables and their relative importance in the ANNs were compared with previous studies to assess its ecological relevance. The results of this study highlight the need to propose mitigation measures to conserve and restore these Mediterranean rivers.

3.2 METHODS

3.2.1 STUDY AREA

The data for this study were collected in the main stream of the Júcar, Cabriel and Turia river basins (Spain), which flows to the Mediterranean Sea (Figure 1). The Spanish Ministry of Environment developed an official ecotypes classification (based on the B system) to implement the Water Framework Directive –WFD– (CEDEX, 2005). The upper reaches of these studied watersheds have slightly regulated or unregulated flows, and they are classified as Mediterranean calcareous mountain rivers (CMM) ecotype. Downstream, large reservoirs strongly regulate the river flows (Alarcón, Contreras and Benagéber reservoirs, Fig. 1). Below these reservoirs, the ecotype is mineralized Mediterranean-continental Rivers (MCM) in the Júcar and Cabriel rivers, whereas we find mineralized rivers of Mediterranean low mountain (ML) and low-altitude Mediterranean rivers (MML) in the Turia River. The region is characterized by a high seasonal and interannual variability of the flow regime; the discharge pattern is strongly dependent on rainfall variability, the driest period occurs in summer and the wet ones are spring and fall (Belmar *et al.,* 2010; Blondel and Aronson, 1999; Gasith and Resh, 1999). The meteorological phenomena known as "*gota fria*" (*medicanes* / *kaltlufttropfen*) may happen in October and November, meaning the occurrence of storms with short duration and high intensity. The predominant soils are highly permeable; this condition implies that the infiltration is an important hydrological process promoting the percolation to aquifer systems (CHJ, 2007). The mean annual temperature ranges between 11.6 and 17 ºC, with the maximum values in July and August (Estrela *et al.,* 2004).

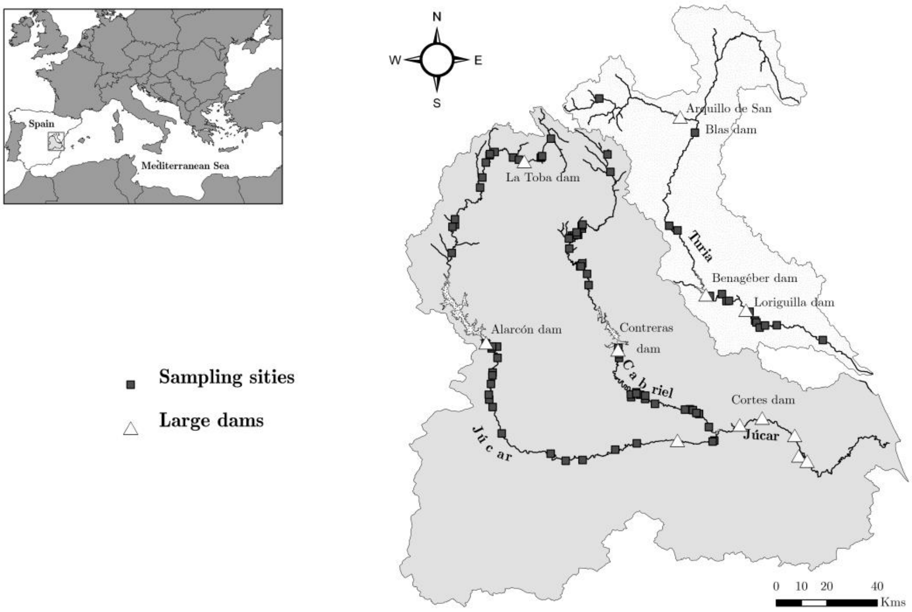

Fig. 1. Location of sampling sites in the Júcar, Cabriel and Turia River Basins

Mediterranean rivers are characterized by a high number of endemic fish species with a reduced range of distribution in contrast to other locations in Europe (Doadrio, 2001; Granado-Lorencio, 1996, 2000; Hermoso and Clavero, 2011). These species are well adapted to the high hydrologic variability of Mediterranean zones (Granado-Lorencio, 2000) and cyprinids is the most abundant family of fish species (Ferreira *et al.*, 2007). In these rivers, we have observed the presence of invasive species, which come from Asia, North America, Europe and other watersheds of the Iberian Peninsula.

3.2.2 PREDICTORS AND RESPONSE VARIABLE

To build the ANN models, we used data from 90 sampling points in the three rivers. The sampling was carried out by electrofishing method in spring and summer, from 2005 to 2009. Each sampling station was kept open, so we did not use nets to close the sampling sites. Each captured individual was classified at species level and returned to the water body. We found 23 fish species in the whole study area, of which 12 were native. Despite this high number of native species, the maximum fish richness was 5 species, this is representative of Mediterranean conditions, in which is frequent to find low values of fish richness per site (Aparicio *et al.*, 2011; Ferreira *et al.*, 2007). According to the International Union for Conservation of Nature (IUCN), 75% of the sampled species are considered as threatened with global extinction (Table 1). This study considers that exotic species and translocated species from other Iberian watersheds are invasive species.

Table 1. Conservation status (according to IUCN, 2012) of fish species sampled in Júcar, Cabriel y Turia Rivers

Species	Family	Native/Nonnative	Threat status
Anguilla anguilla	Anguillidae	Native	CR
Parachondrostoma arrigonis	Cyprinidae	Native	CR
Parachondrostoma turiense	Cyprinidae	Native	EN
Achondrostoma arcasii	Cyprinidae	Native	VU
Barbus haasi	Cyprinidae	Native	VU
Cobitis paludica	Cobitidae	Native	VU
Luciobarbus guiraonis	Cyprinidae	Native	VU
Squalius pyrenaicus	Cyprinidae	Native	NT
Squalius valentinus	Cyprinidae	Native	VU
Iberocypris alburnoides	Cyprinidae	Native	VU
Salmo trutta	Salmonidae	Native	LC
Salaria fluviatilis	Blenniidae	Native	LC
Gobio Gobio	Cyprinidae	Translocated	
Gobio Lozanoi	Cyprinidae	Translocated	
Alburnus alburnus	Cyprinidae	Nonnative	
Cyprinus carpio	Cyprinidae	Nonnative	
Gambusia holbrooki	Poeciliidae	Nonnative	
Esox lucius	Esocidae	Nonnative	
Micropterus salmoides	Centrarchidae	Nonnative	
Sander lucioperca	Percidae	Nonnative	
Oncorhynchus mykiss	Salmonidae	Nonnative	
Lepomis gibbosus	Centrarchidae	Nonnative	
Pseudochondrostoma polylepis	Cyprinidae	Nonnative	

Notes: Key to abbreviations: CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern.

To evaluate the relative effect of invasive species and habitat alteration on native fish species, we used three groups of variables: biological (interactions with fish and aquatic macroinvertebrates availability), habitat, and geographic variables describing the natural gradient from headwater to the mouth (i.e., not subjected to direct human alteration). Some of them were measured in situ, others calculated from geographical information systems (GIS), and others were obtained from the monitoring networks for water quality, biological components and stream flow of the Júcar River Basin Authority (see Table 2). All the biological variables were collected in situ, except the number of macroinvertebrate families, which was supplied by the Júcar River Basin Authority.

Regarding the biological interactions, the introduction of invasive species is one of the most important threats for the conservation of native ichthyofauna, because this threats native fish survival and genetic integrity (Cucherousset and Olden, 2011; Hermoso *et al.,* 2011); specifically, the mechanisms of interaction between invasive and native species may be diverse. Some invasive species can cause severe damage to the habitat by removing materials from the channel bed (Doadrio and Aldeguer, 2007). Others can act as predators of native fish species at different phases of lifecycle (eggs, larval, fry, juvenile, adult) (Granado-Lorencio, 1996), which can result in a reduction of native fish diversity and abundance or the extinction of local species (Bernardo *et al.,* 2003; Crivelli, 1995; Cucherousset and Olden, 2011). Furthermore, native species loss may be related to the reduction of food availability, because macroinvertebrates are the principal source of food for several fish species in Mediterranean rivers (Doadrio and Aldeguer, 2007; Granado-Lorencio, 1996). According to the aforementioned causes of native fish decline, we considered the following biological variables: Invasive fish species richness, number of invasive fish which affect the physical habitat, number of invasive fish predators and number of families of benthic macroinvertebrates (as a surrogate to food availability).

Habitat variables of two categories were considered: water quality and hydromorphology (including connectivity and riparian quality). The use of predictive variables acting at multiple spatial scales usually results in better model performance than single-scale based models (Olden *et al.,* 2006). The multiscale approach allows the integrative analysis of multiple factors and a better understanding of the biodiversity patterns in streams and rivers, in order to support effective conservation and management actions (Filipe *et al.,* 2010). Therefore, we considered different spatial scales based on mesohabitat measurements (i.e., scale of hydro-morphological units, hereafter HMU) and larger scales (e.g. fluvial segment, drainage area).

Water quality is a critical attribute that affects fish species distribution because water pollution alters the survival of fish (Granado-Lorencio, 2000; Jackson *et al.,* 2001). For this reason, we took into account physicochemical properties of water (Table 2). These variables were averaged for the year when fish were sampled; the available data were recorded every 3 months in the official monitoring network of the Júcar River Basin Authority. Moreover, a biologically-based index of water quality (IBMWP) was used. The IBMWP (Alba-Tercedor *et al.*, 2002) ranges from 64 to 260 in our database, it has been widely used for evaluating water quality and for ecological monitoring in Spain (Carballo *et al.*, 2009); this index is based on scores calibrated for species sensitive to water pollution, and it has a very low correlation (ρ^2=0.12) with the number of macroinvertebrates families, thus they are considered independent. These variables were obtained from field data (biomonitoring network) collected by the Júcar River Basin Authority.

Regarding hydromorphological variables, the regime of river discharge (magnitude, pattern of variability) influences the lifecycle traits of Mediterranean fish species (Ferreira *et al.,* 2007; Granado-Lorencio, 2000), thus we considered predictive variables related to flow characteristics. The mean monthly flow was estimated at each sampling location by interpolation between gauging stations, this interpolation was based on the relationship between flow discharge in the natural regime and the accumulated drainage area at each location (Caissie, 2006; Caissie and El-Jabi, 1995; Leopold and Maddock, 1953; Leopold *et al.*, 1964; Olaya-Marín *et al.,* 2012). It was considered that HMUs are linked to the definition of prevailing habitat for fish life at meso and microhabitat level (Bernardo *et al.,* 2003; Costa *et al.,* 2012). We used the classification of HMUs proposed by Dolloff *et al.* (1993), and modified for these Mediterranean rivers as pool, glide, riffle, run and rapid, which was applied in other studies of Mediterranean rivers (Alcaraz-Hernández *et al.*, 2011; Costa *et al.*, 2012); the proportions of HMUs were estimated in field. Transverse hydraulic structures acts as barriers for fish migration, obstructs the colonization of uninhabited reaches and improves the formation of lentic habitats which favour exotic fish species (Granado-Lorencio, 2000); these effects were considered with the use of channel length without artificial barriers (Olaya-Marín *et al.,* 2012). The freshwater fauna is safeguarded by riverine dynamics (Naiman *et* *al.,* 1993; Patten, 1998); the riparian vegetation offers refuge and food, and regulate water channel temperature and the inflow of pollutants from lateral flows (Hattermann *et al.,* 2006; Quinn *et al.,* 2004). To include riverine effects in our models we used the index of riparian habitat quality (Munné *et al.*, 2003) of extensive use in Mediterranean rivers (Garófano-Gómez *et al.,* 2011; Olaya-Marín *et al.*, 2012).

The geographic variables (basin-scale) contribute to describe the natural variability of the river basin (i.e., not subjected to direct human alteration) and are considered as relevant factors to describe the fish assemblages and species richness (Leprieur *et al.*, 2009b; Light and Marchetti, 2007; Oberdorff *et al.*, 1995; Reyjol *et al.*, 2007). Thus we selected watershed area, distance from the headwater source and altitude as potential variables in the models with biological and habitat variables, because they contribute to the models disregarding the effects of invasive species or habitat degradation. The variables supported on geographical information systems were calculated using the ArcGIS ™ 9.3.1 software (ESRI©2009).

Table 2. Biological and habitat variables used as potential predictors. The source of each variable is indicated: in situ, GIS, water quality and flow monitoring networks (MN) or biomonitorinog network (BMN).

Variable	Method	Code	Mean	Range
Biological				
Invasive fish species richness	*In situ*	IFR	2.12	0-5
Number of invasive fish which affect the physical habitat	*In situ*	IFH	0	0-2
Number of invasive fish predators	*In situ*	IFP	1	0-4
Number of families of benthic macroinvertebrates	BMN	BMF	5	4-6
Habitat				
Physicochemical water quality variables:				
Dissolved oxygen (mg/l)	MN	DIS	9.58	8-11
Biological Oxygen Demand (mg/l)	MN	BOD	2.51	2.0-4.0
Total phosphorus (mg/l)	MN	TOP	0.06	0.02-0.22
Nitrites (mg/l)	MN	NIT	0.02	0.01-0.23
pH	MN	PH	8.18	7.77-8.34
Suspended solids (mg/l)	MN	SUS	11.39	3.14-25.21
Conductivity (µS/cm)	MN	CON	797.87	499.16-1210
Water temperature (ºC)	MN	WAT	13.38	5.76-16.75
Iberian monitoring Working Party (IBMWP) index	BMN	IBMWP	131.68	64-260

Hydromorphological				
Channel length without artificial barriers (km)	GIS	CWB	26.35	0.58-95.82
Mean width of water surface (m)	*In situ*	WID	12.46	3.15-19
Mean Annual flow rate (m^3/s)	MN	FMA	4.33	0.03-11.31
Inter-annual mean flow (calculated for 5 years) (m^3/s)	MN	FIA	5.505	0.11-12.36
Coefficient of variation of mean monthly flows (fish sampling year)	MN	FIM	0.58	0.28-0.94
Coefficient of variation of mean annual flows (calculated for 5 years	MN	FCV	0.40	0.15-0.81
Index of riparian habitat quality -QBR	BMN	QBR	73.61	10-100
Hydromorphological units:				
Pools (%)	*In situ*	POO	48.66	0.00-95.00
Glide (%)	*In situ*	GLI	11.21	0.00-80.00
Riffle (%)	*In situ*	RIF	28.27	0.00-89.00
Rapid (%)	*In situ*	RAP	5.79	0.00-53.00
Run (%)	*In situ*	RUN	6.07	0.00-50.00
Geographic (not subjected to direct human alteration)				
Altitude (m a.s.l.)	GIS	ALT	746.48	92.00-1363
Drainage area (km^2)	GIS	DRA	3318.84	54.00-10952
Distance from headwater source (Km)	GIS	DHS	150.19	20.53-327.49

3.2.3 ARTIFICIAL NEURAL NETWORKS MODELLING

It was selected the Multilayer Perceptron (MLP) type of ANN, which applies supervised learning and is the most used in ecology (Özesmi *et al.,* 2006), MLP is able to recognize patterns and to represent complex and nonlinear systems (Lek *et al.,* 2005). The training of this kind of ANN implies the minimization of the error function (e.g. difference between the observed and predicted output) through the iterative modification of weights. A detailed description of MLP can be found at Brosse *et al.* (2003), Goethals *et al.,* (2007) and Olden *et al.,* (2008).

Three ANN models were trained to analyse the effect of habitat alteration and invasive species on native fish richness; the first one was built with biological variables, the second one was made with habitat variables, and the last one was elaborated with both kinds of variables. Before training, the inputs and targets were scaled to fall within a specified range (Demuth *et al.,* 2010); to this end, we used the "*mapminmax*" function of Matlab in the range [-1,+1]. The scaled variables have the same order of magnitude to compute the activation function in the hidden layer (hyperbolic tangent). In this work, we assumed that the inputs of the model with the best performance involve a better control on native species prediction than the inputs of the models with lower performances. Variables were selected by two techniques. In the first step, collinearity among the potential predictors was verified by hierarchical cluster analysis using squared Spearman correlations (ρ^2) as similarity measure. If two variables were correlated with ρ^2 higher than 0.5, in general we used the one with the strongest ecological interpretability (Dormann and Kaschner, 2011; Olaya-Marín *et al.,* 2012). In the second step, we used the forward stepwise method (Gevrey *et al.,* 2003), which allowed us to eliminate irrelevant input variables and reduce the complexity of the ANN architecture (Gevrey *et al.,* 2003; Tirelli and Pessani, 2009).

To define the optimal number of neurons in the hidden layer we tested different ANN architectures. All of these architectures used a single hidden layer because it is parsimonious and sufficient for statistical applications (Bishop, 1996). The algorithm for training the ANN was the Levenberg-Marquardt; this is the fastest method to train neural networks of moderate size (Karul *et al.,* 2000) and recommended by several authors (Gutiérrez-Estrada and Bilton, 2010; Singh *et al.,* 2009; Tan and Van Cauwenberghe, 1999). A detailed description of this algorithm is discussed in Shepherd (1997). The performance of each model was represented by the correlation coefficient (*r*) and the mean square error (MSE). To assess the predictive performance in validation we employed the *k*-fold cross-validation method. The optimal value of *k* was estimated empirically comparing the performance of different ANNs (with *k* from 3 to 10). The contribution of input variables was calculated through the partial derivatives method (PaD) (Dimopoulos et al., 1999; Dimopoulos *et al.,* 1995), which provided two results. First, the profile of the output variations derived from small changes of each predictive variable; the positive values in the PaD plot indicate a positive relationship between the corresponding input variable and the output variable, and vice versa. Secondly, PaD method makes an estimation of the relative contribution of input variables to the prediction of native fish species richness (NFSR).

3.3 RESULTS

According to Figure 2, the variables DHS, WAT, and ALT were strongly correlated with DRA (ρ^2 = 0.77, 0.77 and 0.81 respectively); the last one was selected in the analysis because previous researches have shown its relevance for fish richness (Oberdorff *et al.*, 1995; Olaya-Marín *et al.*, 2012). Therefore, DHS, WAT, and ALT were removed from the dataset. Another correlation (ρ^2 = 0.51) was found between CON and FCV; thus we kept FCV as a potential predictive variable because it represent temporal flow variability (e.g. Olaya-Marín *et al.*, 2012). FIA and FMA were highly correlated (ρ^2 = 0.51), but we preserve both of them because they are both important for Mediterranean endemic freshwater fish (Corbacho and Sánchez, 2001; Costa *et al.*, 2012; Hermoso and Clavero, 2011). IFR had a strong correlation with IFP (ρ^2 = 0.62); however, we used both because they are important for freshwater fish conservation (Hermoso *et al.*, 2011; Smith and Darwall, 2006); thus, in this case the selection of the most significant relied on the forward stepwise method of variables selection.

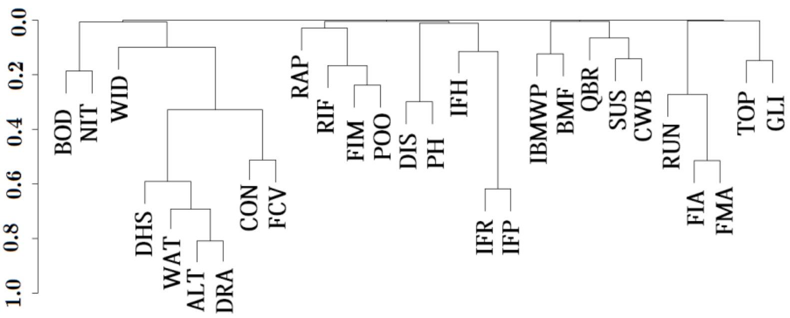

Fig. 2. Hierarchical clustering using squared Spearman correlation (ρ^2) on the environmental variables, in order to indicate their similarities. The variables' codes are noted in table 2.

According to the hierarchical clustering and the forward stepwise procedure, the final variables to predict native fish richness covered the considered spatial scales (i.e. mesohabitat scale and watershed scale). The best model with biological variables had architecture with three layers, five inputs (predictive variables), and four neurons in the hidden layer and one neuron in the output layer (5→4→1). This output layer calculates the predicted value of native fish richness. The best architecture for the model developed with habitat variables, as well as in the third-combined- model was 7→6→1 (Figure 3).

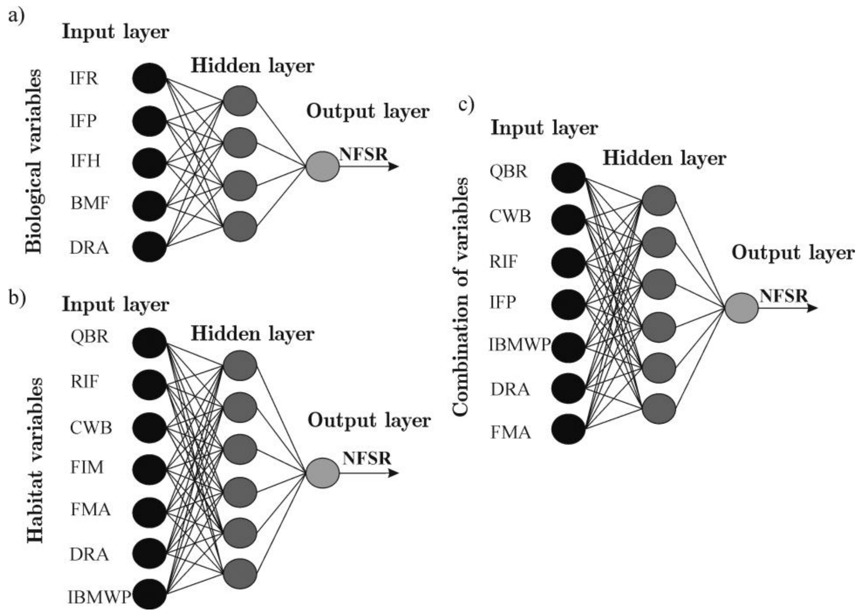

Fig. 3. Artificial neural networks' architecture for the prediction of native fish species richness (NFSR) in a) ANN with biological variables; b) ANN with habitat variables; c) ANN with biological and habitat variables. The variables' codes are noted in table 2.

The model built with habitat variables had a better performance to predict NFSR than the models made with biological variables and the combined model (Table 3). Based on the PaD (Fig. 4), IBMWP and RIF were the most important variables in this model, with a relative importance of 20.72% and 20.18%, respectively. The combined model presented a slightly smaller performance, with these relevant variables: RIF (26.32%) and CWB (20.72%); there was a relatively small contribution (< 15%) of biological variables in this model, where only IFP was present. The performance was considerably smaller in the model with biological variables, where the inputs with the highest weights were: IFP (34.25%) and DRA (27.34%). Figure 4 shows the relative contribution of each variable calculated with the PaD algorithm.

Table 3. Correlation coefficient (r) and mean square error (MSE) between observed and predicted native fish richness in training and validation. Results shows the performance of the three kinds of ANN models

Models	training		Validation	
	r	MSE	*r*	MSE
ANN - biological variables	0.65	1.03	0.67	1.06
ANN - habitat variables	0.90	0.35	0.81	0.62
ANN - both kind of variables	0.88	0.41	0.78	0.64

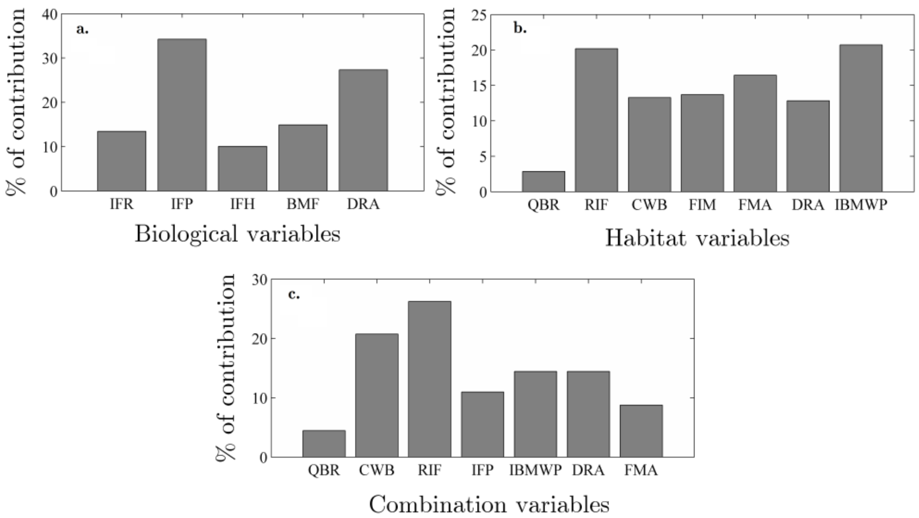

Fig. 4. Relative contribution of each input variable for the prediction of native fish species, according to the partial derivatives algorithm a) model with biological variables b) model with habitat variables c) combined model with biological and habitat variables.

Figure 5 present the profiles for the combined model. This figure shows that the PaD of the native fish richness related to QBR, RIF, CWB, IBMWP and FMA are predominantly positive. Therefore, an increase in those variables leads to increments of native fish species in the study area. On the contrary, the partial derivative values of NFSR with respect to IFP and DRA are predominantly negative.

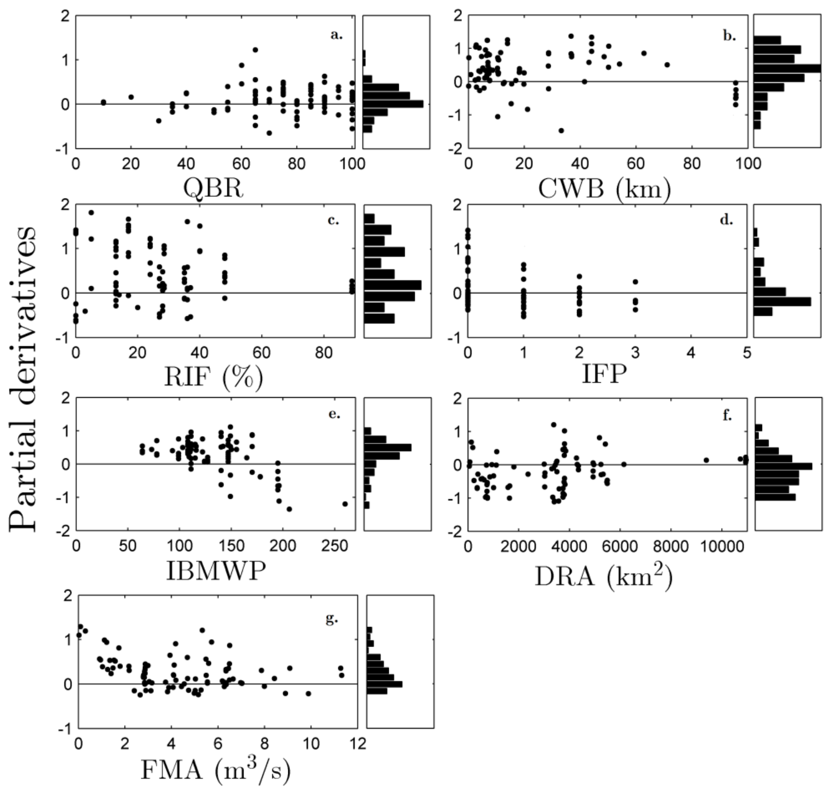

Fig. 5. Partial derivatives of the ANN model response (NFSR) with respect to each independent variable (PaD algorithm, derivatives profile).

3.4 DISCUSSION

According to results, invasive species have a limited importance for the management of the native fish species at the basin scale in the Júcar, Cabriel and Turia rivers. This is evidenced by the facts that the model with habitat variables had the best performance to predict fish richness (Table 3) and the importance of IFP was relatively low in the combined model. One of the advantages of the present work is the integration of biological, habitat and geographic variables in a basin-scale model, with successful results demonstrated with a high performance. The most important variables in the combined model were RIF and CWB, followed by DRA and IBMWP; however, IFP had a smaller importance as the fifth in the ranking of variables importance (Fig. 4c). Considering the two best models, the major importance of RIF was demonstrated. Accordingly, habitat degradation is the main factor affecting the declining of native fish richness in the study area. This result is consistent with the findings of Corbacho and Sánchez (2001), who concluded that habitat degradation could be the main cause in the decline of native fish species in the Guadiana River Basin.

According to literature, the role of invasive species combined with habitat had not been analysed before in the Júcar River Basin. However, some studies in Guadiana river basin (Southwestern Iberian Peninsula) categorized invasive species as drivers of the native fish decline. Godinho and Ferreira (1998) concluded that the presence of invasive species explains the decline of native freshwater fish assemblages in the Guadiana river. Hermoso *et al.* (2011) found that invasive species were the leading driver of the decline of native freshwater fish assemblages in the Guadiana river, considering other conditions like the habitat degradation and the natural gradient of physical factors. Therefore, they proposed a set of measures to prioritize the control of invasive species over all kind of conservation measures, such as eradication or long-term control of invasives, improvement of the river flow regime and reduction of dispersal rates from reservoirs. The discrepancy of our results with the above can be explained by the methodological approach and the ample spectrum of environmental variables considered in the present study.

Firstly, Godinho and Ferreira (1998) and Hermoso *et al.* (2011) used canonical correspondence analysis and structural equation modelling, respectively; these methods are sensitive to certain underlying assumptions in data (e.g. independence, homoscedasticity and normality), which is difficult to satisfy with ecological data. This limitation is relevant to question the validity and reliability of this kind of models (Breiman, 2001; Drew *et al.,* 2011; Guisan and Zimmermann, 2000); the main problem is that functional relationships modelled in SEM are assumed to be linear and therefore is less suitable to situations involving complex relationships between variables (Austin, 2007). Furthermore, we used a method (ANN) able to deal with the inherent variability and non-linearity associated with ecological and biological data; ANNs are better to recognize patterns in data, and generate lesser uncertainty in predictive results than conventional methods (Kang *et al.*, 2011; Lek *et al.*, 2005; Olden *et al.*, 2008). It was previously demonstrated that the relationship between fish community descriptors and environmental variables is very complex and nonlinear in the Júcar, Cabriel and Turia rivers (Olaya-Marín *et al.*, 2012). Secondly, we considered different kind of variables to cover a wide range of dimensions (Table 1): mesohabitat variables (e.g. rifle, pools, run), fluvial connectivity, biological interactions (e.g. invasive fish species richness, number of invasive fish predators, etc.), water quality, hydrology (describing spatial-temporal variability and magnitude of river flow) and geographic variables, which are important for fish communities in Mediterranean environments (Granado-Lorencio, 2000; Olaya-Marín *et al.*, 2012; Poff *et al.*, 1997). Godinho and Ferreira (1998) and Hermoso *et al.* (2011) used a smaller source of potential predictive variables; e.g., they neglected the hydrological variability and the effect of small weirs on river connectivity, which could affect the relative importance of habitat variables in their models. Fish communities are linked to complex processes where biotic and environmental relationships are relevant to explain fish extinction and decline (Granado-Lorencio, 1996, 2000).

Our results demonstrated that habitat variables are key factors of native fish prediction in the studied area. Considering the averaged contribution in the two best models (with similar high performance), the ranking of importance was the following, RIF (23.25%), IBMWP (17.56%), CWB (17.00%), DRA (13.62%), FMA (12.60%) and QBR (3.65%) . Accordingly, the importance of RIF and CWB (Fig. 4c) was previously shown in the Júcar River Basin (Olaya-Marín *et al.*, 2012). The sensitivity analysis points out a positive relationship of RIF and NFSR (Fig. 5c); this is supported by previous studies in Mediterranean rivers, in which RIF was the preferred mesohabitat of native species; by contrast, invasive species lives in lentic habitats (e.g. Bernardo *et al.*, 2003; Olaya-Marín *et al.*, 2012). The preference of riffles is associated to the substrate properties in this mesohabitat, which is adequate for the reproduction of some cyprinids, serve as nursery habitat and refuge (Baras *et al.*, 1996; Bernardo *et al.*, 2003; Martínez-Capel and Garcia de Jalón, 1999; Olaya-Marín *et al.*, 2012).

Regarding the slow-water habitats created by the barriers, invasive species prefer these habitat types (e.g. pools) because they originally come from seasonally stable rivers from a hydrological point of view, unlike Mediterranean systems, which displays a high hydrological variability and a strong seasonality (Belmar *et al.,* 2010; Ferreira *et al.,* 2007; Gasith and Resh, 1999; Vila-Gispert *et al.,* 2005). Discharge fluctuation regulate the structure and dynamics of Mediterranean aquatic communities (Bonada *et al.,* 2007; Magalhães *et al.,* 2007), and Mediterranean fish have adapted their life cycle corresponding to the evolution of their aquatic ecosystems. Granado-Lorencio (1996) highlights that the scarce of lentic natural aquatic systems in the Iberian Peninsula forced ichthiofauna to develop survival strategies (*r* strategies) like: small species with early sexual maturity, short life cycle, and high fecundity (Bernardo *et al.,* 2003; Ferreira *et al.,* 2007; Granado-Lorencio, 2000). Invasive species are known to possess the opposite strategies (Doadrio and Aldeguer, 2007).

The relative contribution by variable revealed that channel length without artificial barriers (CWB) is another key variable for the fish species richness, accordingly with the relation between barriers and slow-water habitats, and with recent works in the region (Alexandre and Almeida, 2010; Corbacho and Sánchez, 2001; Olaya-Marín *et al.,* 2012; Solá *et al.,* 2011). The sensitivity analysis showed a positive relationship between CWB and NFSR (Fig. 5b). Reduced river connectivity is one of the main causes for decline of many continental Iberian fish species (e.g. Aparicio *et al.,* 2000; Solá *et al.,* 2011), dams and weirs disrupt the longitudinal continuity of discharge, hindering and disturbing fish migration (Meixler *et al.,* 2009). These alterations could lead to the extinction of threatened species like *Anguilla anguilla*. Doadrio (2001) found relevant effects of weirs construction on endemic cyprinids, especifically Júcar nase (*Parachondrostoma arrigonis*) and Turia nase (*Parachondrostoma turiense*) because these species need to migrate to the upper parts of the catchments for spawning. Olaya-Marín *et al.* (2012) concluded that artificial barriers (dams and weirs) play a fundamental role in the dramatic reduction of rheophilous native species in Júcar river basin. Native fish richness diminish and invasive species increases when weirs and artificial channels are present (Corbacho and Sánchez, 2001). Therefore, the removal of small barriers, with the consequent improvement in RIF and similar fast-water habitats is a primary measure for the conservation of fish communities. Together with the legal process of water rights cessation, managers have two fundamental tools for river restoration at the basin scale, which should be widely applied to improve the ecological status of rivers. The potential improvements on fish diversity after the removal of obsolete structures have been modelled with ANN in Mediterranean rivers (Olaya-Marín *et al.*, 2012). Furthermore, this is a commonly used method for river enhancement in Europe (Kroes *et al.*, 2006); it is well known that the lack of connectivity in terms of water, sediment, and fauna has important ecological consequences, since the hydromorphological and biological conditions of the ecosystem are directly or indirectly affected (Cowx and Welcomme, 1998). Although the improvement of river connectivity (with fish passes) is an interesting measure to improve the populations, it is not expected that it may produce the same results, given that RIF is not incremented; however, the creation of riffle habitats in nature-like bypasses, as well as spawning channels, provide habitat for a portion of the fish community (Aarestrup *et al.*, 2003; Jormola, 2011; Santos *et al.*, 2005) thus it could partially contribute to the recruitment of fish populations.

Another relevant result (see Figure 5g) was the positive relationship between NFSR and mean annual flow. The FMA magnitude contributes to regulate the dynamic interactions among habitat, floodplain and riparian vegetation (Garófano-Gómez *et al.*, 2012; Olaya-Marín *et al.*, 2012; Poff *et al.*, 1997), and is critical in the interpretation of the spatial distribution of fish Mediterranean communities (Belmar *et al.*, 2011). The hydrological regime in the Iberian Peninsula is strongly altered (Benejam *et al.*, 2010a) and the FMA is expected to reduce with climate change (CEDEX, 2011) in a 10–25%; such scenario would have severe consequences for native fish richness in the Júcar River Basin (Olaya-Marín *et al.*, 2012) and for the structure and dynamics of Mediterranean aquatic communities (Bonada *et al.*, 2007b; Magalhães *et al.*, 2007). The pivotal role of flow regime promotes the conservation of biodiversity, the biotic integrity of lotic systems (Poff *et al.*, 1997; Poff and Zimmerman, 2010), and the growth and survival of native species (Baron and Poff, 2004; Bunn and Arthington, 2002; Poff and Zimmerman, 2010). Flow regime influences water quality, physical features of habitat and energy flows (Baron and Poff, 2004); more specifically, extreme flows (floods and low flows) perform selective pressures (favoring or disfavoring) on the establishment of species (Bunn and Arthington, 2002; Hart and Finelli, 1999).

To maintain dynamical patterns of discharge in the natural range of variation contribute to mitigate or prevent the establishment and proliferation of invasive species; invasive species are unable adapting to the natural Mediterranean hydrological regime (Bernardo *et al.,* 2003; Olaya-Marín *et al.,* 2012; Vila-Gispert *et al.,* 2005). They would need a long time to develop an evolutionary strategy in local hydrological regimes (Doadrio and Aldeguer, 2007). It is therefore urgent for the managers to improve the environmental flow regimes, given that an ongoing extinction crisis is affecting Europe's freshwater fishes (Freyhof and Brooks, 2011). However, the main actual obstacles for the application of environmental flow management are (Richter, 2010): the lack of understanding of environmental flow benefits; uncoordinated management of water resources; low priority given to environmental flows in allocation systems; environmental flow allocations are usually limited to low flows; not addressing the problem of unnatural augmentation of river flows (too much water can be damaging as well); and, difficulty of implementing complex environmental flow specifications.

The human-induced alteration in the magnitude and variability of river discharge influences the habitat of rheophilous species, which commonly are endemic cyprinids threatened with extinction (Freyhof and Brooks, 2011). Invasive predator species like Northern pike (*Esox lucius*), pikeperch (*Sander lucioperca*) and Largemouth bass (*Micropterus salmonides*) are limnophilous, and their lentic habitats are usually favoured by the discharge regime alteration and small environmental flows. Based on our findings, NFSR declines when IFP increases (Fig. 5d); this result is supported by other studies because IFP seriously affects native fish (especially Cyprinidae family). Cyprinids have small size and are easily preyed by invasive species (Doadrio and Aldeguer, 2007), most invasive species exhibit predatory habits and are placed in the top levels in the food web, and Iberian fish rarely are ichthyophagous (Godinho *et al.,* 1997; Granado-Lorencio, 1996). Therefore, invasive predators have an important role in the structure and organization of Mediterranean fish communities, because they can alter the native fish species distribution and diminish its abundance (Doadrio and Aldeguer, 2007; Granado-Lorencio, 1996, 2000). For example, *Micropterus salmoides* is a common predator in the Iberian Peninsula; during the first stages it feeds on macroinvertebrates and plankton, but the adults are ichthiophagous (Doadrio and Aldeguer, 2007; Granado-Lorencio, 1996). Almeida *et al.* (2012) argue that regulated Iberian streams may provide both suitable food and habitat suitability with insignificant predation pressures, thereby may serve as a recruitment environment for *M. salmoides*. Other adverse impacts on native species caused by the introduction of invasive species are the insertion of foreign pathologies, genetic and habitat degradation (Cucherousset and Olden, 2011; García-Berthou, 2001; Granado-Lorencio, 1996).

In these rivers we found that NFSR increases for high values of IBMWP and QBR (Fig. 5e, 5a). These indices have been commonly used by Spanish water authorities to assess biological quality in Mediterranean rivers (MMARM, 2008). IBMWP and QBR are positively related to different aspects of the ecological status of rivers; hence, a good riparian and water quality status contribute to preserve native fish communities (Meador and Goldstein, 2003; Mouton *et al.,* 2012; Naiman *et al.,* 2000; Olaya-Marín *et al.,* 2012; Wichert and Rapport, 1998). Low water quality causes negative effects on metabolism, growth and reproduction of freshwater fish, and morphological anomalies like eroded fins, lesions and tumors (Aparicio *et al.*, 2011; Benejam *et al.*, 2010b). A well-developed riparian cover at riverside provides habitat heterogeneity, food sources, refuge, improvement of water quality, temperature control, bed structure and sediment balance (Mouton *et al.,* 2012; Pinto *et al.,* 2006; Sabater and Tockner, 2010). The linkages between riparian vegetation and fish communities have received little attention in Mediterranean rivers, although riversides have experienced persistent alterations of their natural conditions (Garófano-Gómez *et al.,* 2012; Pardo et al., 2002). Some studies have documented successful examples of how stream flow management can greatly benefit the fish and riparian communities (Rood *et al.,* 2003). On the contrary, Meador and Goldstein (2003) showed that poor riparian conditions are associated with the decreasing of fish communities, this is supported by the fact that fish community structure is strongly controlled by instream habitat (Paller *et* *al.,* 2000). Therefore, among other measures for habitat improvement, the control of water pollution and the riparian management are very important in the fish conservation plans.

In summary, our results support that the ecological status of native fish communities in altered rivers could be improved by the implementation of these prioritized restoration actions: (1) removal of disused weirs to increase the channel length without artificial barriers and reduce the presence of lentic habitats (García de Jalón *et al.,* 2007; Olaya-Marín *et al.,* 2012; WWF, 2009); (2) Wastewater control; (3) application and monitoring of environmental flows imitating the natural hydrological variability, i. e. low flows in dry seasons and flash floods in rainy seasons (Arthington *et al.,* 2006; Costa *et al.,* 2012); (4) sustainable management of vegetation (Meador and Goldstein, 2003); (5) declaration of areas for conservation of native fish species (Filipe *et al.,* 2004; Hermoso *et al.,* 2011); and (6) prevention of new invasive species introductions (Lockwood *et al.,* 2007; Wittenberg and Cock, 2001). However, these prioritized actions must be carefully interpreted in the context of each river basin, where a different ranking of the environmental controls is possible, and the physical habitat hierarchy in the lotic ecosystems must be considered, as well as the scales at which the physical habitat act as a filter on aquatic communities (Poff and Ward, 1990).

ACKNOWLEDGEMENTS

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the projects SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and POTECOL "Evaluación del Potencial Ecológico de Ríos Regulados por Embalses y Desarrollo de Criterios para su mejora según la Directiva Marco del Agua" (CGL2007-66412). We thank to Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment), especially Javier Ferrer and Amparo Piñón, and Junta de Comunidades de Castilla-La Mancha, especially Enrique Montero, for the data provided to develop this study. We thank to Sasa Plestenjak for the collaboration in building the first fish database elaborated in this research, and the colleagues participating in the field work, Rui M.S. Costa, Rafa Muñoz and Virginia Garófano.

REFERENCES

Aarestrup K, Lucas MC, Hansen JA. Efficiency of a nature-like bypass channel for sea trout (Salmo trutta) ascending a small Danish stream studied by PIT telemetry. Ecol. Freshw. Fish 2003; 12: 160-168.

Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, *et al.* Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. BioScience 2008; 58: 403-414.

Alba-Tercedor J, Jáimez-Cuellar P, Álvarez M, Avilés J, Bonada N, Casas J, *et al.* Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP’). Limnetica 2002; 21: 175-185.

Albañez-Lucero MO, Arreguín-Sánchez F. Modelling the spatial distribution of red grouper (Epinephelus morio) at Campeche Bank, México, with respect substrate. Ecol. Model. 2009; 220: 2744-2750.

Alcaraz-Hernández JD, Martínez-Capel F, Peredo-Parada M, Hernández-Mascarell AB. Mesohabitat heterogeneity in four mediterranean streams of the Jucar river basin (Eastern Spain). Limnetica 2011; 30: 363-378.

Alexandre CM, Almeida PR. The impact of small physical obstacles on the structure of freshwater fish assemblages. River Res. Appl. 2010; 26: 977-994.

Almeida D, Almodóvar A, Nicola GG, Elvira B, Grossman GD. Trophic plasticity of invasive juvenile largemouth bass Micropterus salmoides in Iberian streams. Fisheries Research 2012; 113: 153-158.

Aparicio E, Carmona-Catot G, Moyle PB, García-Berthou E. Development and evaluation of a fish-based index to assess biological integrity of Mediterranean streams. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2011; 21: 324-337.

Aparicio E, Vargas MJ, Olmo JM, de Sostoa A. Decline of native freshwater fishes in a Mediterranean watershed on the Iberian Peninsula: A quantitative assessment. Environ. Biol. Fishes 2000; 59: 11-19.

Arthington AH, Bunn SE, Poff NL, Naiman RJ. The challenge of providing environmental flow rules to sustain river ecosystems. Ecol. Appl. 2006; 16: 1311-1318.

Austin M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Model. 2007; 200: 1-19.

Baras E, Philippart JC, Nindaba J. Importance of gravel bars as spawning grounds and nurseries for european running water cyprinids. In Ecohydraulics 2000: 2nd International Symposium on Habitat Hydraulics, vol. A. A. Leclerc, M, Capra, H, Valentin, S, Boudreault, A, Côté, Y, editors. INRS-Eau: Quebec; 1996. p. 367–378.

Baron JS, Poff NL. Sustaining healthy freshwater ecosystems. Water Resources Update 2004; 127: 52-58.

Belmar O, Velasco J, Martinez-Capel F. Hydrological classification of natural flow regimes to support environmental flow assessments in Intensively regulated Mediterranean Rivers, Segura River Basin (Spain). Environ. Manage. 2011; 47: 992-1004.

Belmar O, Velasco J, Martínez-Capel F, Marín AA. Natural flow regime, degree of alteration and environmental flows in the Mula stream (Segura River basin, SE Spain). Limnetica 2010; 29: 353-368.

Benejam L, Angermeier PL, Munné A, García-Berthou E. Assessing effects of water abstraction on fish assemblages in Mediterranean streams. Freshw. Biol. 2010a; 55: 628-642.

Benejam L, Benito J, García-Berthou E. Decreases in Condition and Fecundity of Freshwater Fishes in a Highly Polluted Reservoir. Water, Air, & Soil Pollution 2010b; 210: 231-242.

Bernardo JM, Ilhéu M, Matono P, Costa AM. Interannual variation of fish assemblage structure in a Mediterranean river: implications of streamflow on the dominance of native or exotic species. River Res. Appl. 2003; 19: 521-532.

Bishop CM. Neural Networks for Pattern Recognition. Oxford: University Press; 1996.

Blondel J, Aronson J. Biology and Wildlife of the Mediterranean Region. Oxford: Oxford University Press; 1999.

Bonada N, Rieradevall M, Prat N. Macroinvertebrate community structure and biological traits related to flow permanence in a Mediterranean river network. Hydrobiologia 2007; 589: 91-106.

Breiman L. Statistical modeling: the two cultures. Stat. Sci. 2001; 16: 199-231.

Brosse S, Arbuckle CJ, Townsend CR. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. Biodivers. Conserv. 2003; 12: 2057-2075.

Brosse S, Guegan J-F, Tourenq J-N, Lek S. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 1999; 120: 299-311.

Bulleri F, Balata D, Bertocci I, Tamburello L, Benedetti-Cecchi L. The seaweed Caulerpa racemosa on Mediterranean rocky reefs: from passenger to driver of ecological change. Ecology 2010; 91: 2205-2212.

Bunn SE, Arthington AH. Basic Principles and Ecological Consequences of Altered Flow Regimes for Aquatic Biodiversity. Environ. Manage. 2002; 30: 492-507.

Caissie D. River discharge and channel width relationships for New Brunswick rivers. Canadian Technical Report of Fisheries and Aquatic Sciences, 2637, 2006, pp. 26.

Caissie D, El-Jabi N. Comparison and regionalization of hydrologically based instream flow techniques in Atlantic Canada. Can. J. Civ. Eng. 1995; 22: 235-246.

Carballo R, Cancela J, Iglesias G, Marín A, Neira X, Cuesta T. WFD indicators and definition of the ecological status of rivers. Water Resour. Manag. 2009; 23: 2231-2247.

CEDEX. Caracterización de los tipos de ríos y lagos. Versión 4. Madrid: Centro de Estudios Hidrográficos del CEDEX; 2005.

CEDEX. Evaluación del impacto del cambio climático en los recursos hídricos en régimen natural. Resumen ejecutivo. Available at: http://marm.es/es/agua/formacion/. Madrid: Ministerio de Medio Ambiente y Medio Rural y Marino; CEDEX; 2011.

Clavero M, García-Berthou E. Homogenization dynamics and introduction routes of invasive freshwater fish in the Iberian Peninsula. Ecol. Appl. 2006; 16: 2313-2324.

Corbacho C, Sánchez JM. Patterns of species richness and introduced species in native freshwater fish faunas of a Mediterranean-type basin: the Guadiana River (southwest Iberian Peninsula). Regul. River. 2001; 17: 699-707.

Costa RMS, Martínez-Capel F, Muñoz-Mas R, Alcaraz-Hernández JD, Garófano-Gómez V. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (river Cabriel, Spain). River Res. Appl. 2012; 28: 740-752.

Cowx IG, Welcomme RL. Rehabilitation of Rivers for Fish. Oxford: Wiley; 1998.

Crivelli AJ. Are fish introductions a threat to endemic freshwater fishes in the northern Mediterranean region? Biol. Conserv. 1995; 72: 311-319.

Cucherousset J, Olden JD. Ecological Impacts of non-native freshwater fishes. Fisheries 2011; 36: 215-230.

CHJ. La ictiofauna como elemento de calidad de los ríos de la demarcación hidrográfica del río Júcar. Valencia: Confederación Hidrográfica del Júcar; 2007.

Demuth H, Beale M, Hagan M. Neural Networks toolbox 6. Users Guide. Matlab. Natick, Massachusetts The MathWorks, Inc.; 2010.

Didham RK, Tylianakis JM, Gemmell NJ, Rand TA, Ewers RM. Interactive effects of habitat modification and species invasion on native species decline. Trends Ecol. Evol. 2007; 22: 489-496.

Didham RK, Tylianakis JM, Hutchison MA, Ewers RM, Gemmell NJ. Are invasive species the drivers of ecological change? Trends Ecol. Evol. 2005; 20: 470-474.

Dimopoulos I, Chronopoulos J, Chronopoulou-Sereli A, Lek S. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecol. Model. 1999; 120: 157-165.

Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 1995; 2: 1-4.

Doadrio I. Atlas y Libro Rojo de los Peces Continentales de España. Madrid: Museo Nacional de Ciencias Naturales; 2001.

Doadrio I. Origen y Evolución de la Ictiofauna Continental Española. En:aAtlas y libro rojo de los peces continentales de España. Madrid: CSIC y Ministerio del Medio Ambiente; 2002.

Doadrio I, Aldeguer M. La invasión de especies exóticas en los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Dolloff CA, Hankin DG, Reeves GH. Basinwide estimation of habitat and fish populations in streams: U.S. Department of Agriculture. Forest Service. Southeastern Forest Experiment Station; 1993.

Dormann CF, Kaschner K. Where's the sperm whale? A species distribution example analysis. Supplementary Material to Dormann 2011, Chapter 13, In: Jopp *et al.* 2011, MCED, Springer Verlag.; 2011.

Drew CA, Wiersma Y, Huettmann F. Predictive species and habitat modeling in landscape ecology: concepts and applications. New York: Springer; 2011.

Estrela T, Fidalgo A, Fullana J, Maestu J, Pérez MA, Pujante AM. Júcar Pilot River Basin, provisional article 5 report Pursuant to the Water Framework Directive. Valencia: Confederación Hidrográfica del Júcar; 2004.

Ferreira T, Oliveira J, Caiola N, De Sostoa A, Casals F, Cortes R, *et al.* Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Manag. Ecol. 2007; 14: 473-481.

Filipe AF, Filomena Magalhães M, Collares-Pereira MJ. Native and introduced fish species richness in Mediterranean streams: the role of multiple landscape influences. Divers. Distrib. 2010; 16: 773-785.

Filipe AF, Marques TA, Tiago P, Ribeiro F, Da Costa LM, Cowx IG, *et al.* Selection of priority areas for fish conservation in Guadiana river basin. Conservation Biology 2004; 18: 189-200.

Freyhof J, Brooks E. European Red List of Freshwater Fishes. Luxembourg: Publications Office of the European Union; 2011.

García-Berthou E. Size- and depth-dependent variation in habitat and diet of the common carp (Cyprinus carpio). Aquat. Sci. 2001; 63: 466-476.

García-Berthou E, Alcaraz C, Pou-Rovira Q, Zamora L, Coenders G, Feo C. Introduction pathways and establishment rates of invasive aquatic species in Europe. Can. J. Fish. Aquat. Sci. 2005; 62: 453-463.

García de Jalón D, Sánchez Navarro R, Serrano J. Alteraciones de los regímenes de caudales de los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Garófano-Gómez V, Martínez-Capel F, Bertoldi W, Gurnell A, Estornell J, Segura-Beltrán F. Six decades of changes in the riparian corridor of a Mediterranean river: a synthetic analysis based on historical data sources. Ecohydrology 2012: n/a-n/a.

Garófano-Gómez V, Martínez-Capel F, Peredo-Parada M, Olaya-Marín EJ, Muñoz-Mas R, Costa R, *et al*. Assessing hydromorphological and floristic patterns along a regulated Mediterranean river: The Serpis River (Spain). Limnetica 2011; 30: 307-238.

Garzón MB, Blazek R, Neteler M, Dios RSd, Ollero HS, Furlanello C. Predicting habitat suitability with machine learning models: The potential area of Pinus sylvestris L. in the Iberian Peninsula. Ecol. Model. 2006; 197: 383-393.

Gasith A, Resh VH. Streams in mediterranean climate regions: Abiotic Influences and Biotic Responses to Predictable Seasonal Events. Annu. Rev. Ecol. Syst. 1999; 30: 51-81.

Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 2003; 160: 249-264.

Godinho FN, Ferreira MT. The relative influences of exotic species and environmental factors on an Iberian native fish community. Environ. Biol. Fishes 1998; 51: 41-51.

Godinho FN, Ferreira MT, Cortes RV. The environmental basis of diet variation in pumpkinseed sunfish, Lepomis gibbosus, and largemouth bass, Micropterus salmoides, along an Iberian river basin. Environ. Biol. Fishes 1997; 50: 105-115.

Goethals P, Dedecker A, Gabriels W, Lek S, De Pauw N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 2007; 41: 491-508.

Granado-Lorencio C. Ecología de peces. Sevilla: Universidad de Sevilla; 1996.

Granado-Lorencio C. Ecología de comunidades: el paradigma de los peces de agua dulce. Sevilla: Universidad de Sevilla; 2000.

Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecol. Model. 2000; 135: 147-186.

Gurevitch J, Padilla DK. Are invasive species a major cause of extinctions? Trends Ecol. Evol. 2004; 19: 470-474.

Gutiérrez-Estrada JC, Bilton DT. A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters. Ecol. Model. 2010; 221: 1451-1462.

Hart DD, Finelli CM. Physical-Biological Coupling in Streams: The Pervasive Effects of Flow on Benthic Organisms. Annu. Rev. Ecol. Syst. 1999; 30: 363-395.

Hattermann FF, Krysanova V, Habeck A, Bronstert A. Integrating wetlands and riparian zones in river basin modelling. Ecol. Model. 2006; 199: 379-392.

Hermoso V, Clavero M. Threatening processes and conservation management of endemic freshwater fish in the Mediterranean basin: a review. Mar. Freshwater Res. 2011; 62: 244-254.

Hermoso V, Clavero M, Blanco-Garrido F, Prenda J. Invasive species and habitat degradation in Iberian streams: an analysis of their role in freshwater fish diversity loss. Ecol. Appl. 2011; 21: 175-188.

Hilbert DW, Ostendorf B. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. Ecol. Model. 2001; 146: 311-327.

Ibarra AA, Gevrey M, Park Y-S, Lim P, Lek S. Modelling the factors that influence fish guilds composition using a back-propagation network: assessment of metrics for indices of biotic integrity. Ecol. Model. 2003; 160: 281-290.

Jackson DA, Peres-Neto PR, Olden JD. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 2001; 58: 157–170.

Jormola J. Mitigation, compensation and restoration of habitats in constructed rivers: nature-like bypass channels Proceedings of I Congreso Ibérico de Restauración Fluvial RESTAURARÍOS. León-Spain: Centro Ibérico de Restauración Fluvial, MARM y C. Hidrográfica del Duero; 2011. p. 59-66.

Kang HY, Rule RA, Noble PA. 9.09 - Artificial Neural Network Modeling of Phytoplankton Blooms and its Application to Sampling Sites within the Same Estuary. In: Editors-in-Chief: Eric W, Donald M, editors. Treatise on Estuarine and Coastal Science. Academic Press, Waltham, 2011, pp. 161-172.

Karul C, Soyupak S, Çilesiz AF, Akbay N, Germen E. Case studies on the use of neural networks in eutrophication modeling. Ecol. Model. 2000; 134: 145-152.

Kroes MJ, Gough PP, Wanningen H, Schollema P, Ordeix M, Vesely D. From sea to source. Practical guidance for the restoration of fish migration in European Rivers. Interreg IIIC Project “Community Rivers”. Groningen, The Netherlands; 2006.

Lek S, Scardi M, Verdonschot P, Descy JP, Park YS, editors. Modelling community structure in freshwater ecosystems. Berlin: Springer-Verlag; 2005.

Leopold LB, Maddock T. The hydraulic geometry of stream channels and some physiographic implications. Washington: U.S. Govt. Print. Off.; 1953.

Leopold LB, Wolman MG, Miller JP. Fluvial processes in geomorphology. San Francisco: W.H. Freeman; 1964.

Leprieur F, Olden JD, Lek S, Brosse S. Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. J. Biogeogr. 2009; 36: 1899-1912.

Light T, Marchetti MP. Distinguishing between invasions and habitat changes as drivers of diversity loss among California's freshwater fishes. Conservation Biology 2007; 21: 434-446.

Lippitt CD, Rogan J, Toledano J, Sangermano F, Eastman JR, Mastro V, *et al.* Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. Ecol. Model. 2008; 210: 339-350.

Lockwood JL, Hoopes MF, Marchetti MP. Invasion Ecology. Oxford: Wiley-Blackwell; 2007.

MacDougall AS, Turkington R. Are invasive species the drivers or passengers of change in degraded ecosystems? Ecology 2005; 86: 42-55.

Magalhães MF, Beja P, Schlosser IJ, Collares-Pereira MJ. Effects of multi-year droughts on fish assemblages of seasonally drying Mediterranean streams. Freshw. Biol. 2007; 52: 1494-1510.

Martínez-Capel F, Garcia de Jalón D. Desarrollo de Curvas de preferencia de microhábitat para Leuciscus pyrenaicus y Barbus bocagei por buceo en el río Jarama (Cuenca del Tajo). Limnetica 1999; 17: 71-83.

Mastrorillo S, Dauba F, Oberdorff T, Guégan J-F, Lek S. Predicting local fish species richness in the garonne river basin. Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 1998; 321: 423-428.

Meador MR, Goldstein RM. Assessing Water Quality at Large Geographic Scales: Relations Among Land Use, Water Physicochemistry, Riparian Condition, and Fish Community Structure. Environ. Manage. 2003; 31: 0504-0517.

Meixler MS, Bain MB, Todd Walter M. Predicting barrier passage and habitat suitability for migratory fish species. Ecol. Model. 2009; 220: 2782-2791.

MMARM. Orden MARM/2656/2008 de 10 septiembre, por la que se aprueba la instrucción de planificación hidrológica. BOE núm. 229, de 22 de septiembre de 2008. Madrid: Ministerio de Medio Ambiente, y Medio Rural y Marino (MMARM); 2008.

Mouton AM, Buysse D, Stevens M, van den Neucker T, Coeck J. Evaluation of riparian habitat restoration in a lowland river. River Res. Appl. 2012; 28: 845-857.

Moyle PB. Conservation of native freshwater fishes in the Mediterranean-type climate of California, USA: A review. Biol. Conserv. 1995; 72: 271-279.

Munné A, Prat N, Solà C, Bonada N, Rieradevall M. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2003; 13: 147-163.

Naiman RJ, Bilby RE, Bisson PA. Riparian Ecology and Management in the Pacific Coastal Rain Forest. BioScience 2000; 50: 996-1011.

Naiman RJ, Decamps H, Pollock M. The role of riparian corridors in maintaining regional biodiversity. Ecol. Appl. 1993; 3: 209-212.

Oberdorff T, Guégan J-F, Hugueny B. Global scale patterns of fish species richness in rivers. Ecography 1995; 18: 345-352.

Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernández JD. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). Sci. Total Environ. 2012; 440: 95-105.

Olden JD, Lawler JJ, Poff NL. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Olden JD, Poff NL, Bledsoe BP. Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. Ecol. Inform. 2006; 1: 33-42.

Özesmi SL, Tan CO, Özesmi U. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 2006; 195: 83-93.

Paller MH, Reichert MJM, Dean JM, Seigle JC. Use of fish community data to evaluate restoration success of a riparian stream. Ecological Engineering 2000; 15, Supplement 1: S171-S187.

Pardo I, Álvarez M, Casas J, Moreno JL, Vivas S, Bonada N, *et al*. El hábitat de los ríos mediterráneos. Diseño de un índice de diversidad de hàbitat. Limnetica 2002; 21: 115-132.

Patten D. Riparian ecosytems of semi-arid North America: Diversity and human impacts. Wetlands 1998; 18: 498-512-512.

Penczak T. Fish assemblages composition in a natural, then regulated, stream: A quantitative long-term study. Ecol. Model. 2011; 222: 2103-2118.

Pinto BCT, Araujo FG, Hughes RM. Effects of Landscape and Riparian Condition on a Fish Index of Biotic Integrity in a Large Southeastern Brazil River. Hydrobiologia 2006; 556: 69-83.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, *et al*. The natural klow regime. Bioscience 1997; 47: 769-784.

Poff NL, Ward JV. Physical habitat template of lotic systems: Recovery in the context of historical pattern of spatiotemporal heterogeneity. Environ. Manage. 1990; 14: 629-645.

Poff NL, Zimmerman JKH. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshw. Biol. 2010; 55: 194-205.

Quinn JM, Boothroyd IKG, Smith BJ. Riparian buffers mitigate effects of pine plantation logging on New Zealand streams: 2. Invertebrate communities. For. Ecol. Manage. 2004; 191: 129-146.

Reyjol Y, Hugueny B, Pont D, Bianco PG, Beier U, Caiola N, *et al.* Patterns in species richness and endemism of European freshwater fish. Glob. Ecol. Biogeogr. 2007; 16: 65-75.

Ribeiro F, Leunda PM. Non-native fish impacts on Mediterranean freshwater ecosystems: current knowledge and research needs. Fisheries Manag. Ecol. 2012; 19: 142-156.

Richter BD. Re-thinking environmental flows: from allocations and reserves to sustainability boundaries. River Res. Appl. 2010; 26: 1052-1063.

Rood SB, Gourley CR, Ammon EM, Heki LG, Klotz JR, Morrison ML, *et al.* Flows for floodplain forests: A successful riparian restoration. Bioscience 2003; 53: 647-656.

Sabater S, Tockner K. Effects of Hydrologic Alterations on the Ecological Quality of River Ecosystems Water Scarcity in the Mediterranean. In: Sabater S, Barceló D, editors. 8. Springer Berlin / Heidelberg, 2010, pp. 15-39.

Santos JM, Ferreira MT, Godinho FN, Bochechas J. Efficacy of a nature-like bypass channel in a Portuguese lowland river. Journal of Applied Ichthyology 2005; 21: 381-388.

Shepherd AJ. Second-Order Methods for Neural Networks. New York: Springer-Verlag; 1997.

Singh KP, Basant A, Malik A, Jain G. Artificial neural network modeling of the river water quality--A case study. Ecol. Model. 2009; 220: 888-895.

Smith KG, Darwall WRT, editors. The status and distribution of freshwater fish endemic to the mediterranean basin. Gland, Switzerland/Cambridge, UK.: IUCN -The World Conservation Union; 2006.

Solá C, Ordeix M, Pou-Rovira Q, Sellarés N, Queralt A, Bardina M, *et al.* Longitudinal connectivity in hydromorphological quality assessments of rivers. The ICF index: A river connectivity index and its application to Catalan rivers. Limnetica 2011; 30: 273-292.

Spieles DJ. Protected Land: Disturbance, Stress, and American Ecosystem Management Springer 2010.

Tan Y, Van Cauwenberghe A. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intell. 1999; 12: 21-35.

Tirelli T, Pessani D. Use of decision tree and artificial neural network approaches to model presence/absence of *Telestes muticellus* in piedmont (North-Western Italy). River Res. Appl. 2009; 25: 1001-1012.

Vila-Gispert A, Alcaraz C, García-Berthou E. Life-history traits of invasive fish in small Mediterranean streams. Biol. Invasions 2005; 7: 107-116-116.

Wichert GA, Rapport DJ. Fish Community Structure as a Measure of Degradation and Rehabilitation of Riparian Systems in an Agricultural Drainage Basin. Environ. Manage. 1998; 22: 425-443-443.

Wittenberg R, Cock MJ. Invasive Alien Species: A Toolkit of Best Prevention and Management Practices. Wallingford, Oxon UK: CAB Internacional; 2001.

WWF. Liberando ríos Propuestas de WWF para el desmantelamiento de presas en España. Madrid: WWF España; 2009.

Chapter 4

A comparison of artificial neural networks and random forests to predict native fish species richness in Mediterranean rivers

Esther Julia Olaya-Marín, Francisco Martínez-Capel & Paolo Vezza

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

ABSTRACT

Machine learning (ML) techniques have become important to support decision making in management and conservation of freshwater aquatic ecosystems. Given the large number of ML techniques and to improve the understanding of ML utility in ecology, it is necessary to perform comparative studies of these techniques as a preparatory analysis for future model applications. In this context, the aims of this study were (i) to compare the reliability and ecological relevance of two predictive models for fish richness, based on the techniques of artificial neural networks

(ANN) and random forest (RF) and (ii) to evaluate the concordance in terms of selected important variables between the two modelling approaches. The model performances were evaluated through three performance metrics: the determination coefficient (R^2), the Mean Square Error (MSE) and the adjusted determination coefficient (R^2_{adj}) and both models were developed using a *k*-fold cross validation procedure. According to the results, RF obtained the best performance in training and it was the model with the smallest number of inputs (five predictive variables), while the ANN model required seven input variables. In the cross-validation procedure both techniques gave similar results (R^2=68% for RF and R^2=66% for ANN). Although the two methods selected different subsets of input variables, both of them demonstrated high ecological relevance for the conservation of native fish in the Mediterranean area. This work shows how the use of different modelling methods can assist the critical analysis of predictions and underlines the possible use of ML techniques to design environmental management actions at a catchment scale.

Keywords: Artificial Neural Networks, Random Forests, machine learning, native fish, species richness, Mediterranean rivers.

4.1 INTRODUCTION

In the last decades, due to the worldwide accelerated degradation of freshwater ecosystems (Beechie *et al.*, 2010; Postel, 2000; Strayer and Dudgeon, 2010) ecological modelling has become an important tool for wildlife and habitat conservation (Breckling *et al.,* 2011; Drew *et al.,* 2011). Particularly in Mediterranean rivers, pollution, introduction of exotic species and alteration of hydrological regimes are factors which have influenced fish population decrease and, in some cases, the extinction of native species (Didham *et al.*, 2007; García-Berthou *et al.,* 2005; Smith and Darwall, 2006). According to IUCN, the 56% of freshwater Mediterranean species are threatened (Smith and Darwall, 2006) and, given the high degree of endemicity of biota and its high vulnerability to habitat alteration, more research is currently needed on local and native fish populations (Corbacho and Sánchez, 2001; Doadrio, 2002).

The conservation of fish diversity is one of the most critical issues facing the preservation of Mediterranean biodiversity (Smith and Darwall, 2006); and, due to its sensitivity to human disturbances, fish species richness is widely used as a primary indicator of the ecological changes and as a criterion for the selection of conservation areas (He *et al.,* 2010; Lek *et al.,* 2005; van Jaarsveld *et al.,* 1998). Increasing knowledge about the relationships between environmental features and fish populations is therefore essential for the design of effective habitat conservation and river restoration actions.

Ecological and biological data rarely satisfy the principles of parametric approaches, in which data must be independent, normal and homoscedastic (Breiman, 2001b; Guisan and Zimmermann, 2000). This circumstance increases the challenges in modelling ecological phenomena. To cope with these issues, machine learning (ML) techniques have been widely used due to their ability to drive non-linearity and generate less uncertain predictive results (Olden *et al.,* 2008; Recknagel, 2001).

Several researchers have applied ML in ecological studies (Aertsen *et al.,* 2010; Armitage and Ober, 2010; D'Heygere *et al.,* 2006; Leclere *et al.,* 2011; Mouton *et al.,* 2011). In particular, artificial neural networks (ANN) and random forest (RF) are two machine learning techniques which are currently valuable tools for ecological modelling, especially useful in analysing large datasets and identify non-linear relationships (Drew *et al.,* 2011). ANN are recognized as powerful and effective tools (Mastrorillo *et al.,* 1998; Olden *et al.,* 2008) to solve complex dependencies which are difficult with other traditional statistical methods (Lek *et al.,* 2005; Olden *et al.,* 2008). In the context of freshwater fish studies, ANN have been used with satisfactory results (e.g. Carpenter *et al.,* 1999; Olaya-Marín *et al.,* 2012; Tirelli *et al.,* 2009). Ibarra *et al.* (2003) used ANN and multiple regression models (MLR) to identify the factors that influence fish guilds in the Garonne river basin (south-western France). They found better predictions of fish guilds with ANN than MLR. A similar result about ANN prediction accuracy is reported in

Tirelli and Pessani (2009, 2011), who used ANN and decision trees to predict the presence of *Telestes muticellus* and *Alburnus alburnus alborella* in Piedmont rivers (north-western Italy). Moreover, Tirelli *et al.* (2009) applied ANN, discriminant function analysis, logistic regression and decision tree to model *Salmo marmoratus* distribution in Piedmont (Italy) and the performance of ANN was superior to the other modelling techniques. Also for unbalanced data, Hauser-Davis *et al.* (2010) concluded that ANN are an excellent alternative in classification problems.

Regarding RF, it is currently considered a promising technique in ecology (Cutler *et al.,* 2007; Cheng *et al.,* 2012; Drew *et al.,* 2011; Franklin, 2010) but it has rarely been applied in freshwater fish studies. RF has the ability to identifying and exploring non-intuitive relationships (Evans *et al.,* 2011), with high accuracy and flexibility to perform both regression and classification analyses (Cheng *et al.,* 2012). He *et al.* (2010) compared the use of classification and regression trees (CART) and RF to predict endemic fish assemblages and species richness in the upper Yangtze River. The study showed that RF is better than CART in terms of accuracy and efficiency. Knudby *et al.* (2010) used linear (LM) and generalized additive models (GAM), Bagging, RF, Boosted Regression Trees (BRT) and support vector machines (SVM) to build predictive mapping of reef fish species richness, diversity and biomass. They found that the tree-based models were generally superior to predict species richness of reef fish. Furthermore, Mouton *et al.* (2011) found similar predictive performance of RF and Fuzzy logic models to represent mesohabitat suitability for *Salmo trutta* in Spain, while Kampichler *et al.,* (2010) compared different ML techniques (including ANN and RF) for classification problems and recommend the use of RF in conservation biology.

Given the large number of ML techniques, there is not a protocol to define the most indicated method to address a particular ecological question or management action for freshwater ecosystems. It is therefore necessary to conduct comparative studies of ML techniques for model identification and selection (Guisan and Zimmermann, 2000). Moreover, the knowledge related to the ecology of Mediterranean rivers needs to be deepen and further efforts to improve the understanding of the main factors influencing species richness are valuable (Aparicio *et al.,* 2011; Filipe *et al.,* 2010) . In this context, the aims of this study were (i) to compare the reliability and ecological relevance of two predictive models for fish richness, based on the techniques of ANN and RF and (ii) to evaluate the concordance in terms of selected important variables between the two modelling approaches. It is important to highlight here that a comparison between ANN and RF for prediction of fish species richness has not yet been presented in literature. These comparisons are currently considered a new open line of research (Aertsen *et al.,* 2011) and this paper represents a further contribution in such a field.

4.2 MATERIALS AND METHODS

4.2.1 STUDY AREA AND DATA COLLECTION

This study was carried out with data collected in the main streams of the Júcar, Cabriel and Turia Rivers, in the Eastern Iberian Peninsula (Fig.1). These rivers are characterized by a Mediterranean climate, a flow regime controlled by rainfall variability, a strong seasonal and inter-annual discharge variation with two high flow periods per year (spring and fall) and severe droughts in summer (Blondel and Aronson, 1999; Ollero *et al.*, 2011). The maximum temperatures are registered in July and August, coinciding with the dry period (Estrela *et al.*, 2004) and the mean temperature ranges between 11.6 to 17 ºC. The mean annual precipitation in the study area is 500 mm, ranging between 320 mm in dry years to 800 mm in the wet ones (Estrela *et al.*, 2004). The soils are highly permeable and are characterized by high infiltration and percolation rates (CHJ, 1997; Estrela *et al.*, 2004). During the last decades, the natural flow regime has been altered after the building of reservoirs and water abstractions; flow regulation is severe particularly for streams located in the middle and lower part of the watersheds. The effect of flow regulation is expressed by an inversion of the intra-annual variability pattern; in summer, the regulated flow is greater than natural flow, and in contrast, the regulated flow is smaller than the natural flow during winter (Aparicio *et al.*, 2011). Due to industrial and urban waste water, pollution also affects rivers (Aparicio *et al.,* 2011; Estrela *et al.,* 2004) and agricultural practices, particularly in spring and summer, constitute a source of diffusive pollutants at the catchment-scale (Estrela *et al.*, 2004).

In the analyses, we used data from 90 sampling sites along the main streams of the three rivers (Fig. 1). The sites were selected as representative in terms of river morphology and proportion of mesohabitats which characterize the analysed water courses. Native fish species richness (i.e. the number of fish species at each sampling site) was defined by means of a single-pass electrofishing during the spring/summer period from 2005 to 2009. The limits of the sampling sites were open and the minimum length of each sampled reach was 50 m.

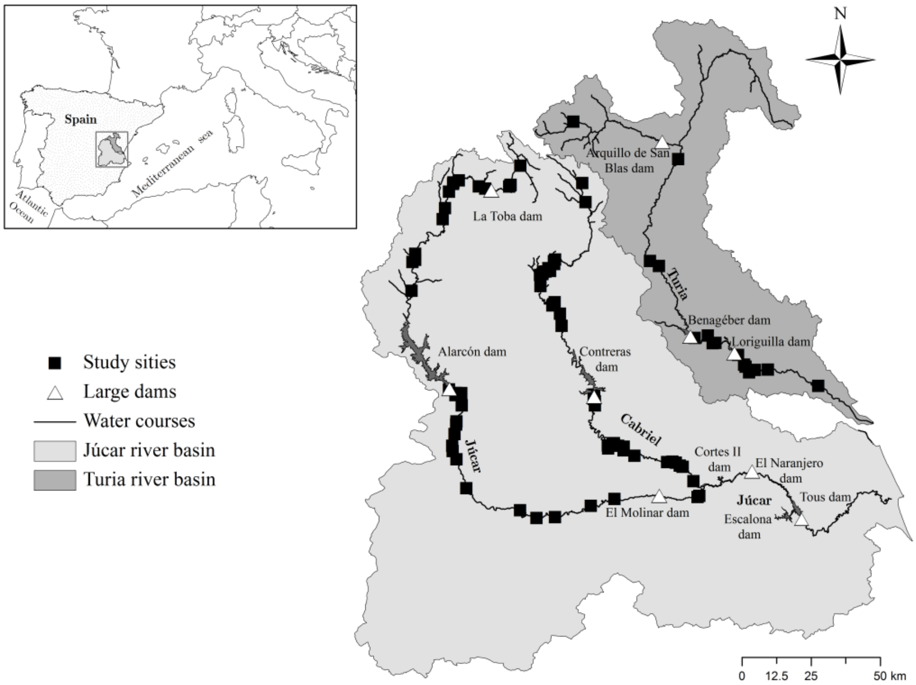

Fig.1. Study area showing the distribution of the 90 sampling sites in the three considered rivers (Jucar, Cabriel and Turia rivers).

The total fish diversity comprises 12 native species (Table 1) with a maximum local richness of 5 species. These values are common in Mediterranean rivers, which are generally characterized by a low species richness per site (Ferreira *et al.*, 2007). Cyprinidae is the predominant family in the three rivers; the most important genera are *Achondrostoma, Parachondrostoma, Luciobarbus, Barbus, Squalius* and *Iberocypris.* Other species present in the rivers are *Cobitis paludica* and *Salaria fluviatilis* which are very sensitive to pollution and have strict environmental requirements (CHJ, 2007). All these species perform small-scale migrations for reproduction within the river system and the only one migrating at large scale is *Anguilla Anguilla,* a catadromous fish species with a complex life-history that includes migrations across the Atlantic Ocean. The number of individuals of these native fish species decreased consistently in the last decades as a consequence of habitat modifications (including barriers) and pollution in the lower river reaches (Costa *et al.*, 2012; CHJ, 2007; Doadrio, 2001a).

Table 1. *Freshwater fish species present in the study area related to its threat status (Freyhof and Brooks, 2011; IUCN, 2012).CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern.*

Species name	Common name	Family	Threat status
Anguilla anguilla	European eel	Anguillidae	CR
Parachondrostoma arrigonis	Júcar nase	Cyprinidae	CR
Parachondrostoma turiense	Turia nase	Cyprinidae	EN
Achondrostoma arcasii	Bermejuela	Cyprinidae	VU
Barbus haasi	Iberian redfin barbell	Cyprinidae	VU
Cobitis paludica	Southern Iberian spined-loach	Cobitidae	VU
Luciobarbus guiraonis	Eastern Iberian barbell	Cyprinidae	VU
Squalius pyrenaicus	Southern Iberian chub	Cyprinidae	NT
Squalius valentinus	Eastern Iberian chub	Cyprinidae	VU
Iberocypris alburnoides	Calandino	Cyprinidae	VU
Salmo trutta	Brown trout	Salmonidae	LC
Salaria fluviatilis	Freshwater blenny	Blenniidae	LC

The environmental variables used in the construction of the ANN and RF models were 24, which were selected considering their ecological importance for fish life cycle (Bernardo *et al.*, 2003; Costa *et al.*, 2012; Granado-Lorencio, 1996;

Jackson *et al.,* 2001; Oberdorff *et al.,* 1995). Variables were obtained from three main sources: in situ, from GIS analyses and through the monitoring network (MN) of the Júcar River Basin (Table 2). Physico-chemical variables (i.e. dissolved oxygen, biological oxygen demand, total phosphorous, nitrite, pH, suspended solids, water temperature) corresponded to the mean annual value referred to the year of the survey. The proportions of hydro-morphological units (HMUs) and the mean width of water surface were measured *in situ*. The classification of HMUs was based on the method proposed by Dollof *et al.* (1993) and five different types were selected: pool, glide, riffle, rapid and run (Alcaraz-Hernández *et al.,* 2011; Costa *et al.,* 2012). Geographical variables (i.e. channel length without artificial barriers, altitude, drainage area and distance from the source) were obtained with the ArcGIS 9.3.1 software (ESRI©2009).

The mean monthly flow was calculated at ungauged sites through a linear interpolation between gauged sites. To define the hydrological indexes (inter-annual mean flow and the coefficients of variation of mean monthly flow and mean annual flow), we used the relationship between the flow in natural conditions and the accumulated drainage area, and then we transformed monthly flow values to regulated conditions (Caissie, 2006a; Caissie and El-Jabi, 1995; Leopold *et al.,* 1964). The riparian habitat quality index (QBR, Munné *et al.,* 2003) was taken into account to assess the morphological conditions of the sampling sites; this index was adopted by the Spanish Ministry of Environment (MMARM, 2008). QBR consists of four components, which synthesize qualitative features related to the conservation state of the riparian area: total vegetation cover, vegetation cover structure and quality, and river channel alterations. The values of this index are distributed in five quality intervals (≥95: excellent quality; 90-75: good quality; 70-55: moderate quality; 30-50: poor quality; ≤ 25: bad quality). Finally, we used the Iberian Biomonitoring Working Party index –IBMWP- (Alba-Tercedor, 1996; Alba-Tercedor and Sánchez-Ortega, 1988) based on invertebrate analysis to evaluate the biological quality of the rivers. IBMWP values are distributed in five ranges of water quality: > 150: very clean water; 101-120: unpolluted water or not appreciably altered; 61-100: partially polluted water with some evident effects; 36-60: polluted water; 16-35: very polluted water; < 15: heavily polluted water.

Table 2. Potential environmental variables used to build the predictive models for native fish species richness. Physico-chemical and hydrological parameters were obtained from the monitoring network (MN) of the Júcar River Basin Authority, mean width and hydro-morphological unit proportions were measured in situ during fish samplings, while geographical variables were derived from GIS analyses.

Variable	**Code**	**Source**	**Mean**	**Standard deviation**
Dissolved Oxygen (mg/l)	DIS	MN	9.58	0.44
Biological oxygen demand (mg/l)	BOD	MN	2.51	0.77
Total phosphorus (mg/l)	TOP	MN	0.06	0.03
Nitrite (mg/l)	NIT	MN	0.02	0.02
pH	PH	MN	8.18	0.11
Suspended solids (mg/l)	SUS	MN	11.39	5.77
Water Conductivity (µS/cm)	CON	MN	797.87	172.62
Water temperature (ºC)	WAT	MN	13.38	2.48
Percentage of pools (%)	POO	*in situ*	48.66	21.42
Percentage of glides (%)	GLI	*in situ*	11.21	16.89
Percentage of riffles (%)	RIF	*in situ*	28.27	21.41
Percentage of rapids (%)	RAP	*in situ*	5.79	6.85
Percentage of runs (%)	RUN	*in situ*	6.07	12.10
Mean width of the water surface (m)	WID	*in situ*	12.46	4.68
Channel length without artificial barriers (km)	CWB	GIS	26.35	29.46
Altitude (m a.s.l)	ALT	GIS	746.48	298.43
Drainage area (km^2)	DRA	GIS	3318.84	2607.51
Distance from the source (km)	DHS	GIS	150.19	76.41
Mean Annual flow rate (m^3/s)	FMA	MN	4.33	2.44
Inter-annual mean flow (calculated for 5 years) (m^3/s)	FIA	MN	5.50	2.57
Coefficient of variation of mean monthly flows (referred to fish sampling)	FIM	MN	0.58	0.18
Coefficient of variation of mean annual flows (calculated for 5 years)	FCV	MN	0.40	0.17
Index of Riparian Habitat Quality	QBR	MN	73.61	20.74
Iberian Biomonitoring Working Party	IBMWP	MN	131.68	36.32

4.2.2 MODELLING TECHNIQUES

4.2.2.1 ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial neural networks are mathematical models inspired in the structure and behaviour of the human brain (Olden *et al.,* 2008). They are considered as a powerful computational tool to address ecological issues that are difficult to analyse by traditional statistical methods (Lek *et al.,* 2005); among different types of ANN, Multilayer Perceptron (MLP) is the most used in ecology (Özesmi *et al.,* 2006). It is constituted by multiple layers and the information is transferred from the input layer to the output one (feed-forward). This kind of ANN has supervised learning, which implies the use of input and output datasets to iteratively change the weights until the simulated outputs are similar to the observed ones. To minimize the error, the algorithm employs the values of the error calculated in the previous iteration and then updates the weights. A detailed description of ANN is reported in Olden *et al.* (2008) and Goethals *et al.* (2007).

This study applied a MLP to predict native fish species richness. For each ANN model, we built and tested several MLP models to establish, by trial and error estimates, the optimal number of neurons in the hidden layer. A single hidden layer was used to significantly reduce the computational time. Moreover, as reported in Kurková (1992) the use of a single hidden layer produces similar results compared to the incorporation of additional hidden layers. To work with values characterized by the same order of magnitude, data of all variables were scaled. The transfer function in the hidden layer was a hyperbolic tangent and an identity function in the case of the output layer. The hyperbolic tangent gave optimal results in previous studies (Isa *et al.*, 2010) in which a performance comparison was carried out to select the best MLP activation function (Olaya-Marín *et al.,* 2012). The Levenberg-Marquardt (LM) optimization algorithm was used to train the candidate models because this algorithm is the fastest method to train neural networks of moderate size (Karul *et al.,* 2000). LM has been applied successfully in ecology (Gutiérrez-Estrada and Bilton, 2010; Tan and Van Cauwenberghe, 1999) and a description of the algorithm can be found in Singh *et al.* (2009). To test and validate the models, we used the *k*-fold cross-validation procedure and tried different values of *k* (ranging from 3 to 10, Dormann, 2011; Goethals *et al.,* 2007). The best *k* value was then identified by the comparison of the performance of the different ANN obtained in the cross-validation procedure. All numerical calculations were performed using MATLAB software (version R2010a).

4.2.2.2 RANDOM FORESTS

The fish richness was also predicted using Random Forest (RF) methodology (Breiman, 2001a; Cutler *et al.,* 2007) in the statistical software R (R Development Core Team, 2009) by means of the randomForest package (Liaw and Wiener, 2002). RF is an ensemble learning technique based on a combination of a large set of decision trees. Each tree is trained by selecting random bootstrap subsets X_i (i = bootstrap iteration which ranges from 1 to t) of the original dataset X and a random set of predictive variables. This aspect represents the main difference compared to standard decision trees (Breiman *et al.,* 1984), where each node is split using the best split among all predictive variables (e.g. Vezza *et al.,* 2012). Moreover, RF corrects many of the known issues in CART, such as over-fitting (Breiman, 2001a; Cutler *et al.,* 2007), and provides very well-supported predictions with large numbers of independent variables (Cutler *et al.,* 2007).

As the response variable (fish richness) was numerical, we confined our attention to regression RF. The algorithm for growing a random forest of t regression trees performed as follows (for full details see Breiman, 2001a):

(1) t bootstrap samples X_i of the training dataset were randomly drawn with replacement, each one containing approximately two third of the elements of the original dataset X. The elements not included in each training dataset are referred to as out-of-bag data (OOB, i.e. the validation dataset) for that bootstrap sample. On average, each element of X was an OOB element in one-third of the t iterations.

(2) For each bootstrap sample X_i, an unpruned regression tree was grown. At each node *m* variables were randomly selected and the best split was automatically chosen.

(3) New data (OOB elements) were predicted by averaging the predictions of the generated *t* trees. In particular, for each element (y_i) of the original dataset an aggregated prediction (g_{OOB}) was developed and the out-of-bag estimate of the error rate (E_{OOB}) was computed as $\left[E_{OOB} = (1/t)\cdot\sum_{1}^{t}\left[y_i - g_{OOB}\left(X_i\right)\right]^2\right]$.

The E_{OOB} was also used to choose an optimal value of *t* and *m* (Breiman, 2001a). As E_{OOB} is an unbiased estimate of the generalization error, it is not necessary to test the predictive ability of the model using a cross-validation procedure (Breiman, 2001a). However, in accordance with ANN and for a more reliable comparison, we performed *k*-fold cross-validation (with *k* ranging from 3 to 10) following the approach reported in Svetnik *et al.* (2003).

4.2.2.3 INPUT SELECTION

As a first step, a correlation matrix was calculated to verify collinearity. For high correlations (Spearman's $rho^2>0.5$) we removed the variable with less ecological relevance (Dormann, 2011). To identify the most important predictive variables we followed two different approaches. On one hand, the forward stepwise procedure was applied in the ANN models; this method consists of adding step by step a single input variable and then evaluating the improvement in ANN performance (Gevrey *et al.,* 2003). The irrelevant input variables are therefore eliminated measuring the complexity reduction of the ANN model (see Gevrey *et al.,* 2003; Tirelli and Pessani, 2009) and at the end of the process the variables that imply a significant improvement in the ANN performances are selected (Fig. 2).

On the other hand, we applied the Model Improvement Ratio technique (MIR, Murphy *et al.,* 2010) to identify the most parsimonious RF model. RF produces a measure of variable importance by analysing the deterioration of the predictive ability of the model when each predictor is replaced in turn by random noise. The increase in the mean square error of each tree (IncMSE) is used as a score of importance of a given variable in regression RF models (Vincenzi *et al.*, 2011), as it indicates the contribution to RF prediction accuracy for that variable. The MIR technique uses the variable importance standardized from zero to one and the improvement ratio was therefore calculated as [*In/Imax*], where *In* is the importance of a given variable and *Imax* is the maximum model improvement score. We then iterated through MIR thresholds (i.e. 0.05 increments), with all variables above the threshold retained for each model (Evans and Cushman, 2009). The models corresponding to different subsets were then compared and the one that exhibits the minimum MSE error and the lowest number of variables was selected.

4.2.2.4 MODEL EVALUATION

The overall accuracy of the two statistical models was evaluated using three performance metrics, which are commonly used in ecological modelling (Aertsen *et al.,* 2011; Chenard and Caissie, 2008; Singh *et al.,* 2009): the determination coefficient (R^2), the Mean Square Error (MSE) and the adjusted determination coefficient (R^2_{adj}).

The determination coefficient (R^2) assesses the proportion of variability explained by the model, it is calculated by:

$$R^2 = \left(\frac{\sum\left(Y^{sim} \cdot Y^{obs}\right) - \left(\left(\sum Y^{sim} \cdot \sum Y^{obs}\right)/n\right)}{\sqrt{\left(\sum Y^{sim^2} - \left(\left(\sum Y^{sim}\right)^2/n\right)\right) \cdot \left(\sum Y^{obs^2} - \left(\left(\sum Y^{obs}\right)^2/n\right)\right)}} \right)^2 \quad (1)$$

Where, Y^{obs} are the observed values, Y^{sim} represent the predicted values, and n is the total number of observations. The Mean Square Error (MSE) is the error between model predictions and observed values, it is expressed as:

$$MSE = \frac{1}{n}\sum\left(Y^{sim} - Y^{obs}\right)^2 \quad (2)$$

The adjusted determination coefficient is a modification of the determination coefficient and was used during the model selection procedures to compare models with different numbers of predictive variables (Vezza *et al.*, 2010). In contrast to R^2, this coefficient penalizes the excessive use of inputs, and it is expressed as follows:

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} \tag{3}$$

where p represents the number of input variables.

Finally, the ecological interpretation of each optimal ANN and RF model was carried out by the assessment of the relative importance of the inputs. For ANN the partial derivatives (PaD) method was applied (Dimopoulos *et al.*, 1999; Dimopoulos *et al.*, 1995), which represents the mostly used approach to evaluate the relative importance in MLP (Gevrey *et al.*, 2003), whereas for RF the relative importance was indicated by the IncMSE values of each variable (Breiman, 2001a).

4.3 RESULTS

The best neural network architecture to predict native fish richness had three layers (i.e. 7→6→1), with seven neurons in the input layer (which corresponds to the predictive variables), a hidden layer with six neurons, and the output layer with a single neuron; the last one calculates the estimated values of native fish richness. During the k-fold cross validation, the ANN performance did not increase with k values higher than 6, then we used k = 6 to validate the model. The stepwise selection of variables and the MSE is illustrated in Fig. 2.

For RF, the out-of-bag estimates of the error rate (E_{OOB}) were used to select the optimum Random Forest parameters m and t, while RF performance estimates were based on a 6-fold cross-validation (Svetnik *et al.*, 2003). The m parameter (number of variables permutated at each node) was defined as [1/3·(number of variables)]

(Breiman, 2001a) and it was 2 in the selected model (see Breiman, 2001a). Moreover, in our analysis the OOB error stabilization occurred between *t* = 1500 and *t* = 2500 replicates. However, a heuristic estimation of *t* taking into account the OOB error stabilization was defined as [2·(*t* for E_{OOB} stabilization) = 5000] (Evans and Cushman, 2009a). The relation between the MSE and number of variables in RF is shown in Fig. 2. For both models, the MSE quickly decreased as the number of input variables was increasing (Fig. 2). A breakpoint was located at 7 variables in ANN; and at 5 variables in RF. Based on this criterion, we used 7 predictive input variables to build the ANN model and 5 for RF.

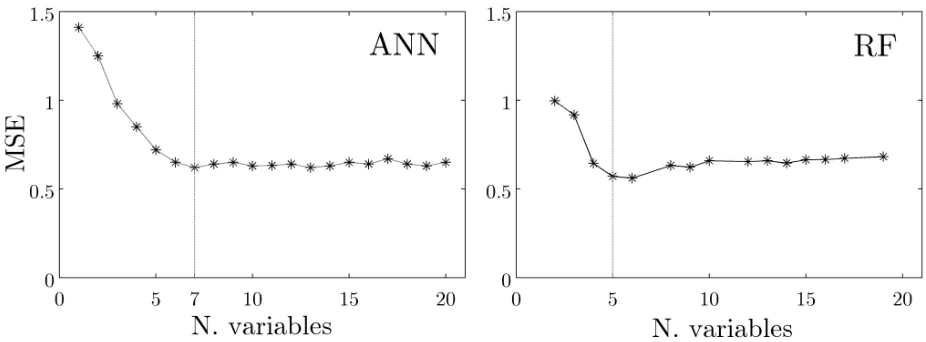

Fig. 2. Artificial Neural Networks (ANN) and Random Forests (RF) performance in terms of Mean Squared Error (MSE) as a function of the number of input variables (N. Variables). The final ANN model (including 7 variables) and RF model (including 5 variables) were those in which the incorporation of any additional variable meant no relevant error decrease (vertical lines).

According to the correlation analysis and the forward stepwise procedure the relevant variables to predict the native fish richness with the ANN model are reported, in order of importance, in Fig. 3: the Iberian Biomonitoring Working Party (IBMWP), percentage of riffles (RIF), mean annual flow rate (FMA), coefficient of variation of mean monthly flow (FIM), channel length without artificial barriers (CWB), drainage area (DRA) and index of riparian habitat quality (QBR). In contrast for RF, the selected variables were QBR, percentage of runs (RUN), DRA, IBMWP and percentage of rapids (RAP).

The best performance in training was obtained by the RF model (R^2 = 0.94; MSE = 0.10), whereas in validation both techniques gave similar results (MSE = 0.62, R^2=66% for ANN; MSE = 0.56, R^2 = 68% for RF). Table 3 displays the performance indices of RF and ANN models.

Table 3. Artificial Neural Networks (ANN) and Random Forest (RF) performance indices in training and k-fold cross validation procedures. Models were evaluated with determination coefficient (R^2), mean squared error (MSE) and adjusted determination coefficient (R^2adj).

Models	Training			*k*=6 Cross Validation		
	R^2	MSE	R^2_{adj}	R^2	MSE	R^2_{adj}
ANN	0.81	0.35	0.78	0.66	0.62	0.63
RF	0.94	0.10	0.94	0.68	0.56	0.66

The implementation of the partial derivatives algorithm for ANN revealed that the most important variables to predict native fish richness were IBMWP, with a relative importance of 20.72%, and percentage of riffle (RIF) with an importance of 20.18%. In the case of the RF model, the most important variables were QBR and percentage of runs (RUN), with a relative importance of 23.51% and 22.02%, respectively (Fig. 3).

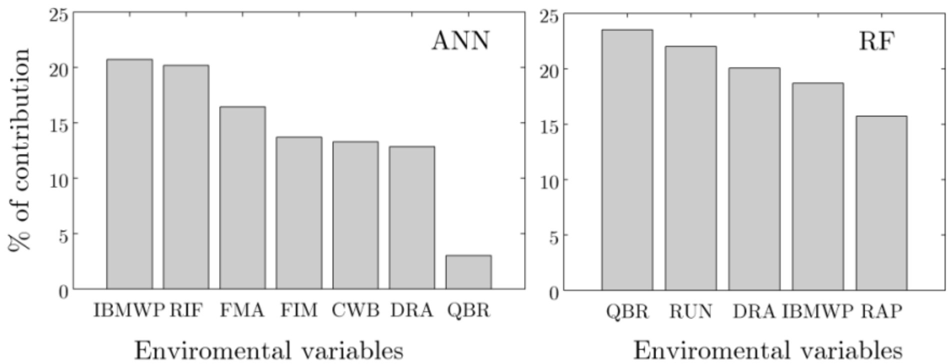

Fig. 3. Relative importance (expressed in % of contribution) of each input variable to predict native fish richness. Left side: the ANN model, right side: the RF model. See codes of variables in Table 2.

4.4 DISCUSSION

In this study two machine learning techniques (i.e. Artificial Neural Networks and Random Forest) were applied to estimate the fish species richness in the Jucar River Basin, as preparatory reference for future wildlife and habitat conservation actions. The methodology compared the reliability and ecological relevance of the two statistical techniques in order to evaluate their applicability and assess the concordance in terms of variables importance between the two predictive models. Looking at the results, ANN and RF showed no significant differences of performance in the cross-validation procedure (R^2=68% for RF and R^2=66% for ANN, Table 3). However, it is important to note that RF outperformed ANN in terms of MSE, particularly considering small numbers of input variables (Fig. 2, N. Variables < 7).

RF was the model with the smallest number of inputs and only five variables were required for prediction, while ANN required seven. Parsimonious models are particularly suitable for applications (fewer variables to be surveyed) and the difference in the number of inputs highlighted the advantage of using RF. Moreover, models with fewer variables are much easier to interpret and can reduce the level of prediction uncertainty (Jorgensen and Fath, 2011). However, the RF model showed much higher accuracy in the calibration to that obtained in the validation phase (Table 3), presenting a considerable difference in performance. In contrast, for ANN the difference between training and validation prediction error was smaller and demonstrated more stable results.

Since the 90's, diverse mathematical algorithms have emerged in order to quantify and interpret the importance and contribution of input variables to the model output, and, at same time, to identify and eliminate redundant variables to increase model parsimony (e.g. Gevrey *et al.*, 2003; Murphy *et al.*, 2010; Olden and Jackson, 2002). In this research we used partial derivatives (Dimopoulos *et al.*, 1999; Dimopoulos *et al.*, 1995) for ANN and the model improvement ratio (Murphy *et al.*, 2010) for RF. PaD allowed the classification of the input variables according to their contribution to the output variable and, in accordance with

Gevrey *et al.* (2003), the technique produced a stable variables ranking over the different ANN models. On the other hand, MIR demonstrated to be a simple and powerful methodology to select the threshold that minimized both retained inputs and model error.

For variable selection, a forward stepwise methodology was embedded in the ANN algorithm, whereas the variable importance values were used in RF to screen the overall range of inputs and select the most parsimonious model. The two procedures were based on different approaches and led to two different sets of variables. However, this result is not surprising and it is confirmed in several studies (Abrahamsson *et al.*, 2003; Reunanen, 2003; Wells *et al.*, 2011; Xu and Zhang, 2001), in which different variable selection procedures produced similar subsets of variables. It is important to note that the RF ranking of variables is based on all possible combinations of model inputs with *m* random variables permuted at each node of the tree. In contrast, the one-step-ahead search procedure of ANN may not lead to the best combination of inputs; it required the modeller to study the sequence of variables and analyse whether the addition or removal of a few more variables might not produce any improvement.

Another important aspect in ecological modelling involves the evaluation and interpretation of the results. The presented models were in accordance with the literature due to fact that the selected input variables, such as water quality, flow regime and the status of riparian forest, are of great importance for the Mediterranean fish populations (Bernardo *et al.*, 2003; Ferreira *et al.*, 2007; García de Jalón *et al.*, 2007; Granado-Lorencio, 1996, 2000).

Both models had in common three variables to predict fish richness: drainage area (DRA), quality index of the riparian forest (QBR) and the biological index for water quality (IBMWP). Although the variables' ranking was not the same (in terms of % of contribution, Fig. 3), this accordance can indicate the robustness of the results (e.g. Xu and Zhang, 2001). In several studies (Leprieur *et al.*, 2009; Oberdorff *et al.*, 1995; Reyjol *et al.*, 2007) DRA is considered an important environmental variable for fish species richness, integrating information related to habitat diversity, the variety of microclimates and flow regimes in the river basin (e.g. Allan and Castillo, 2007; Begon *et al.*, 2006; Townsend *et al.*, 2008), while QBR and IBMWP indicate the water quality and riparian forest key role in determining the richness of native fish. QBR and IBMWP have been already mentioned in Spain as relevant factors influencing fish species richness (Carballo *et al.*, 2009; Sánchez-Montoya *et al.*, 2010) and are widely used for the ecological monitoring of rivers. Indeed, the riparian forest provide shelter and food for aquatic organisms (Naiman *et al.*, 1993) and can strongly influence the quality of aquatic habitats, particularly along a gradient of river regulation (Garófano-Gómez *et al.*, 2011). Furthermore, river pollution is currently one of the most important threats for the Mediterranean freshwater fish (Smith and Darwall, 2006); it can severely disrupt the functioning of the aquatic ecosystem and compromise the survival of biota (Granado-Lorencio, 2000).

Both ANN and RF selected the proportion of HMUs as important predictors of fish richness and, in particular, percentage of riffles (RIF) were selected in ANN and percentage of runs (RUN) and rapids (RAP) in RF. Although the two statistical techniques considered different HMU types, one can observe how the spatial distribution and dynamics of mesohabitats can be of great importance for fish conservation (Fausch *et al.*, 2002). According to Bernardo *et al.* (2003) riffles and runs can influence the composition of Mediterranean fish communities; particularly for those dominated by the family Cyprinidae (Ferreira *et al.*, 2007; Granado-Lorencio, 2000), because these mesohabitats can offer good conditions in terms of food availability and shelters.

The variables related to flow regime (mean annual flow rate, FMA, and the coefficient of variation of mean monthly flow, FIM) along with the channel length without artificial barriers (CWB) were only selected by ANN. Different studies highlighted the role of flow variability and magnitude in supporting the aquatic communities (Belmar *et al.*, 2011; Olaya-Marín *et al.*, 2012; Poff *et al.*, 1997) and, the longitudinal connectivity has important consequences on the distribution of native fish (Kroes *et al.*, 2006), either small or large-scale migratory species (e.g. *Parachondrostoma arrigonis* or *Anguilla Anguilla*, both critically endangered in the region of interest). Including these aspects in fish richness prediction underline the ecological relevance of the ANN model, which seemed to capture the interplay between natural and anthropogenic factors influencing fish species distribution.

As reported in Siroky (2009) RF models are fast to train and tune. In our research the time needed for RF model construction was much smaller than for ANN (few minutes compared to hours) due to the structure of RF algorithm characterized by few parameters to set and a limited number of variables (*m*) to be permutated at each tree node. ANN needed more time for computer architecture design and learning as it performed a large amount of trials changing the number of neurons and the type of activation function in the hidden layer. However, the amount of time needed can be reduced by using different computer processors working in parallel (Armitage and Ober, 2010) and once calibrated, ANN are able to process a large volume of data and quickly generate predictions (Olden *et al.*, 2008).

The applied ML techniques involve elegant mathematical theories and are known to be robust to noise and able to manage the non-linearity among variables (Lek *et al.,* 2005; Olden *et al.,* 2008; Siroky, 2009), but for some authors they can be seen as black boxes (Hooten, 2011). In particular, the relationships between the input variables and the predicted values produced by ANN and RF do not have simple representations such as a formula (e.g. linear regressions) or pictorial graph (e.g. regression trees) that characterizes the entire function, and this lack of simple representation can make the ecological application difficult (Cutler *et al.,* 2007). Olden *et al.* (2002b) provided an interesting point of view to give light into the so called “black box”; but, compared to traditional statistical methods, ML remain more complex to understand and apply (Olden and Jackson, 2002; Olden *et al.,* 2008). In addition, these techniques require the modeller to have knowledge and experience in designing and programming the algorithms in order to make effective use of the tools and to reach satisfactory and valid results (Franklin, 2010).

Although a comprehensive evaluation of several different techniques was beyond the scope of this paper, we believe that the best predictive ML method cannot be chosen *a priori* and both ANN and RF constitute valuable tools to predict fish richness in the Mediterranean area. Looking at the results, we can state that the use of more than one ML technique on the same study area was helpful, not only to identify the most important variables, but also to interpret the goodness and coherence of the results. As operational procedure for future studies on fish species richness, we can state that the comparison of different ML methods should be carried out and, as a further step already planned for the near future, these analyses can be performed in other Mediterranean basins. The presented approaches, which relate environmental variables to the fish communities, can be indeed used for predicting fish richness at the basin scale and can be incorporated into the decision-making process for water resources management (Paredes-Arquiola *et al.*, 2012 in press). For instance they could contribute to perform large-scale assessments of environmental flow standards, based on methodological frameworks with a regional perspective (Paredes-Arquiola *et al.*, 2011 in press; Poff *et al.*, 2010).

ACKNOWLEDGEMENTS

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the projects SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and POTECOL "Evaluación del Potencial Ecológico de Ríos Regulados por Embalses y Desarrollo de Criterios para su mejora según la Directiva Marco del Agua" (CGL2007-66412). In addition, the present research was developed in the frame of the EU-funded HolRiverMed project (Marie Curie Actions and in particular the Intra-European Fellowships for the mobility of European researchers). We thank the Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment) for the data provided to develop this study and we also owe our gratitude to Sasa Plestenjak for the collaboration in building the first fish database for this research.

REFERENCES

Abrahamsson C, Johansson J, Sparén A, Lindgren F. Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets. Chemometrics Intell. Lab. Syst. 2003; 69: 3-12.

Aertsen W, Kint V, Van Orshoven J, Muys B. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). Environ. Modell. Softw. 2011; 26: 929-937.

Aertsen W, Kint V, van Orshoven J, Özkan K, Muys B. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Model. 2010; 221: 1119-1130.

Alba-Tercedor A. Macroinvertebrados acuaticos y calidad de las aguas de los ríos. Almería: IV Simposio del Agua en Andalucía (SIAGA); 1996.

Alba-Tercedor J, Sánchez-Ortega A. Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de Hellawell (1978). Limnetica 1988; 4: 51-56.

Alcaraz-Hernández JD, Martínez-Capel F, Peredo-Parada M, Hernández-Mascarell AB. Mesohabitat heterogeneity in four mediterranean streams of the Jucar river basin (Eastern Spain). Limnetica 2011; 30: 363-378.

Allan JD, Castillo MM. Stream ecology: structure and function of running waters. Netherlands: Springer; 2007.

Aparicio E, Carmona-Catot G, Moyle PB, García-Berthou E. Development and evaluation of a fish-based index to assess biological integrity of Mediterranean streams. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2011; 21: 324-337.

Armitage DW, Ober HK. A comparison of supervised learning techniques in the classification of bat echolocation calls. Ecol. Inform. 2010; 5: 465-473.

Beechie TJ, Sear DA, Olden JD, Pess GR, Buffington JM, Moir H, *et al.* Process-based principles for restoring river ecosystems. Bioscience 2010; 60: 209-222.

Begon M, Townsend C, Harper J. Ecology: From individuals to Ecosystems. Oxford: Blackwell Publishing; 2006.

Belmar O, Velasco J, Martinez-Capel F. Hydrological classification of natural flow regimes to support environmental flow assessments in Intensively regulated Mediterranean Rivers, Segura River Basin (Spain). Environ. Manage. 2011; 47: 992-1004.

Bernardo JM, Ilhéu M, Matono P, Costa AM. Interannual variation of fish assemblage structure in a Mediterranean river: implications of streamflow on the dominance of native or exotic species. River Res. Appl. 2003; 19: 521-532.

Blondel J, Aronson J. Biology and Wildlife of the Mediterranean Region. Oxford: Oxford University Press; 1999.

Breckling B, Jopp F, Reuter H. Backgrounds and scope of ecological modelling: Between intellectual adventure and scientific routine. In: Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, teachers and scientists. Springer-Verlag, Berlin, 2011, pp. 3-12.

Breiman L. Random Forests. Mach. Learn. 2001a; 45: 5-32.

Breiman L. Statistical modeling: the two cultures. Stat. Sci. 2001b; 16: 199-231.

Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Belmont, California: Wadsworth International Group; 1984.

Caissie D. River discharge and channel width relationships for New Brunswick rivers. Canadian Technical Report of Fisheries and Aquatic Sciences, 2637, 2006, pp. 26.

Caissie D, El-Jabi N. Comparison and regionalization of hydrologically based instream flow techniques in Atlantic Canada. Can. J. Civ. Eng. 1995; 22: 235-246.

Carballo R, Cancela J, Iglesias G, Marín A, Neira X, Cuesta T. WFD indicators and definition of the ecological status of rivers. Water Resour. Manag. 2009; 23: 2231-2247.

Carpenter GA, Gopal S, Macomber S, Martens S, Woodcock CE, Franklin J. A Neural Network Method for Efficient Vegetation Mapping. Remote Sens. Environ. 1999; 70: 326-338.

Corbacho C, Sánchez JM. Patterns of species richness and introduced species in native freshwater fish faunas of a Mediterranean-type basin: the Guadiana River (southwest Iberian Peninsula). Regul. River. 2001; 17: 699-707.

Costa RMS, Martínez-Capel F, Muñoz-Mas R, Alcaraz-Hernández JD, Garófano-Gómez V. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (river Cabriel, Spain). River Res. Appl. 2012; 28: 740-752.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, *et al.* Random Forests for classification in ecology. Ecology 2007; 88: 2783-2792.

Chenard J-F, Caissie D. Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. Hydrol. Process. 2008; 22: 3361-3372.

Cheng L, Lek S, Lek-Ang S, Li Z. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 2012; 42: 127-136.

CHJ. Plan hidrológico cuenca del Júcar: sistemas de explotación. Valencia: Confederación Hidrográfica del Júcar; 1997.

CHJ. Estudio general sobre la Demarcación Hidrográfica del Júcar. Madrid: Confederación Hidrográfica del Júcar; 2007.

D'Heygere T, Goethals PLM, De Pauw N. Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. Ecol. Model. 2006; 195: 20-29.

Didham RK, Tylianakis JM, Gemmell NJ, Rand TA, Ewers RM. Interactive effects of habitat modification and species invasion on native species decline. Trends Ecol. Evol. 2007; 22: 489-496.

Dimopoulos I, Chronopoulos J, Chronopoulou-Sereli A, Lek S. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecol. Model. 1999; 120: 157-165.

Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 1995; 2: 1-4.

Doadrio I. Atlas y Libro Rojo de los Peces Continentales de Espana; 2001.

Doadrio I. Origen y Evolución de la Ictiofauna Continental Española. En:aAtlas y libro rojo de los peces continentales de España. Madrid: CSIC y Ministerio del Medio Ambiente; 2002.

Dolloff CA, Hankin DG, Reeves GH. Basinwide estimation of habitat and fish populations in streams: U.S. Department of Agriculture. Forest Service. Southeastern Forest Experiment Station; 1993.

Dormann CF. Modelling species' distributions. In: Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, teachers and scientists. Springer-Verlag, Berlin, 2011, pp. 179-196.

Drew CA, Wiersma Y, Huettmann F. Predictive species and habitat modeling in landscape ecology: concepts and applications. New York: Springer; 2011.

Estrela T, Fidalgo A, Fullana J, Maestu J, Pérez MA, Pujante AM. Júcar Pilot River Basin, provisional article 5 report Pursuant to the Water Framework Directive. Valencia: Confederación Hidrográfica del Júcar; 2004.

Evans J, Cushman S. Gradient modeling of conifer species using random forests. Landsc. Ecol. 2009; 24: 673-683.

Evans JS, Murphy MA, Holden ZA, Cushman SA. Modeling species distribution and change using Random Forest. In: Drew CA, Wiersma YF, Huettmann F, editors. Predictive Species and Habitat Modeling in Landscape Ecology. Springer New York, 2011, pp. 139-159.

Fausch K, Torgersen C, Baxter C, Li H. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. Bioscience 2002; 52: 483-498.

Ferreira T, Oliveira J, Caiola N, De Sostoa A, Casals F, Cortes R, *et al.* Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Manag. Ecol. 2007; 14: 473-481.

Filipe AF, Filomena Magalhães M, Collares-Pereira MJ. Native and introduced fish species richness in Mediterranean streams: the role of multiple landscape influences. Divers. Distrib. 2010; 16: 773-785.

Franklin J. Mapping species distributions: spatial inference and prediction. New York: Cambridge University Press; 2010.

Freyhof J, Brooks E. European Red List of Freshwater Fishes. Luxembourg: Publications Office of the European Union; 2011.

García-Berthou E, Alcaraz C, Pou-Rovira Q, Zamora L, Coenders G, Feo C. Introduction pathways and establishment rates of invasive aquatic species in Europe. Can. J. Fish. Aquat. Sci. 2005; 62: 453-463.

García de Jalón D, Sánchez Navarro R, Serrano J. Alteraciones de los regímenes de caudales de los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Garófano-Gómez V, Martínez-Capel F, Peredo-Parada M, Olaya-Marín EJ, Muñoz-Mas R, Costa R, *et al.* Assessing hydromorphological and floristic patterns along a regulated Mediterranean river: The Serpis River (Spain). Limnetica 2011; 30: 307-238.

Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 2003; 160: 249-264.

Goethals P, Dedecker A, Gabriels W, Lek S, De Pauw N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 2007; 41: 491-508.

Granado-Lorencio C. Ecología de peces. Sevilla: Universidad de Sevilla; 1996.

Granado-Lorencio C. Ecología de comunidades: el paradigma de los peces de agua dulce. Sevilla: Universidad de Sevilla; 2000.

Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecol. Model. 2000; 135: 147-186.

Gutiérrez-Estrada JC, Bilton DT. A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters. Ecol. Model. 2010; 221: 1451-1462.

Hauser-Davis RA, Oliveira TF, Silveira AM, Silva TB, Ziolli RL. Case study: Comparing the use of nonlinear discriminating analysis and Artificial Neural Networks in the classification of three fish species: acaras (Geophagus brasiliensis), tilapias (Tilapia rendalli) and mullets (Mugil liza). Ecol. Inform. 2010; 5: 474-478.

He Y, Wang J, Lek-Ang S, Lek S. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. Sci. Total Environ. 2010; 408: 4211-4220.

Hooten MB. The state of spatial and spatio-temporal statistical modeling. In: Drew C, Wiersma Y, Huettmann F, editors. Predictive Species and Habitat Modeling in Landscape Ecology. Springer New York, 2011, pp. 29-41.

Ibarra AA, Gevrey M, Park Y-S, Lim P, Lek S. Modelling the factors that influence fish guilds composition using a back-propagation network: assessment of metrics for indices of biotic integrity. Ecol. Model. 2003; 160: 281-290.

Isa IS, Omar S, Saad Z, Osman MK. Performance comparison of different multilayer perceptron network activation functions in automated weather classification. Proceedings of the 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation. Kota Kinabalu, Malaysia; 2010. p. 71-75.

IUCN. IUCN Red List of Threatened Species. Version 2012.1;www.iucnredlist.org (downloaded on 14 September 2012), 2012.

Jackson DA, Peres-Neto PR, Olden JD. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 2001; 58: 157–170.

Jorgensen SE, Fath BD. Fundamentals of ecological modelling: applications in environmental management and research. 4th ed. Amsterdam: Elsevier; 2011.

Kampichler C, Wieland R, Calmé S, Weissenberger H, Arriaga-Weiss S. Classification in conservation biology: a comparison of five machine-learning methods. Ecol. Inform. 2010; 5: 441-450.

Karul C, Soyupak S, Çilesiz AF, Akbay N, Germen E. Case studies on the use of neural networks in eutrophication modeling. Ecol. Model. 2000; 134: 145-152.

Knudby A, LeDrew E, Brenning A. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. Remote Sens. Environ. 2010; 114: 1230-1241.

Kroes MJ, Gough PP, Wanningen H, Schollema P, Ordeix M, Vesely D. From sea to source. Practical guidance for the restoration of fish migration in European Rivers. Interreg IIIC Project “Community Rivers”. Groningen, The Netherlands; 2006.

Kurková V. Kolmogorov's theorem and multilayer neural networks. Neural Netw. 1992; 5: 501-506.

Leclere J, Oberdorff T, Belliard J, Leprieur F. A comparison of modeling techniques to predict juvenile 0+ fish species occurrences in a large river system. Ecol. Inform. 2011; 6: 276-285.

Lek S, Scardi M, Verdonschot P, Descy JP, Park YS, editors. Modelling community structure in freshwater ecosystems. Berlin: Springer-Verlag; 2005.

Leopold LB, Wolman MG, Miller JP. Fluvial processes in geomorphology. San Francisco: W.H. Freeman; 1964.

Leprieur F, Brosse S, García-Berthou E, Oberdorff T, Olden JD, Townsend CR. Scientific uncertainty and the assessment of risks posed by non-native freshwater fishes. Fish. Fish. 2009; 10: 88-97.

Liaw A, Wiener M. Classification and regression by Random Forest. R News 2002; 2: 18-22.

Mastrorillo S, Dauba F, Oberdorff T, Guégan J-F, Lek S. Predicting local fish species richness in the garonne river basin. Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 1998; 321: 423-428.

MMARM. Orden MARM/2656/2008 de 10 septiembre, por la que se aprueba la instrucción de planificación hidrológica. BOE núm. 229, de 22 de septiembre de 2008. Madrid: Ministerio de Medio Ambiente, y Medio Rural y Marino (MMARM); 2008.

Mouton AM, Alcaraz-Hernández JD, De Baets B, Goethals PLM, Martínez-Capel F. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Modell. Softw. 2011; 26: 615-622.

Munné A, Prat N, Solà C, Bonada N, Rieradevall M. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2003; 13: 147-163.

Murphy MA, Evans JS, Storfer A. Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology 2010; 91: 252-261.

Naiman RJ, Decamps H, Pollock M. The role of riparian corridors in maintaining regional biodiversity. Ecol. Appl. 1993; 3: 209-212.

Oberdorff T, Guégan J-F, Hugueny B. Global scale patterns of fish species richness in rivers. Ecography 1995; 18: 345-352.

Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernández JD. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). Sci. Total Environ. 2012; 440: 95-105.

Olden JD, Jackson DA. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 2002; 154: 135-150.

Olden JD, Lawler JJ, Poff NL. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Ollero A, Ibisate A, Gonzalo L, Acín V, Ballarín D, Díaz E, *et al.* The IHG index for hydromorphological quality assessment of rivers and streams: updated version Limnetica 2011; 30: 255-262.

Özesmi SL, Tan CO, Özesmi U. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 2006; 195: 83-93.

Paredes-Arquiola J, Martinez-Capel F, Solera A, Aguilella V. Implementing environmental flows in complex water resources systems–case study: the Duero river basin, Spain. River Res. Appl. 2011 in press.

Paredes-Arquiola J, Solera-Solera A, Martínez-Capel F, Momblanch-Benavent A, Andreu-Álvarez J. Integrating water management, habitat modelling and water quality at basin scale environmental flow assessment - Tormes River (Spain). Hydrol. Sci. J.-J. Sci. Hydrol. 2012 in press.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, *et al.* The natural klow regime. Bioscience 1997; 47: 769-784.

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, *et al.* The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshw. Biol. 2010; 55: 147-170.

Postel SL. Entering an era of water scarcity: the challenges ahead. Ecol. Appl. 2000; 10: 941-948.

R Development Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria; 2009.

Recknagel F. Applications of machine learning to ecological modelling. Ecol. Model. 2001; 146: 303-310.

Reunanen J. Overfitting in making comparisons between variable selection methods. J. Mach. Learn. Res. 2003; 3: 1371-1382.

Reyjol Y, Hugueny B, Pont D, Bianco PG, Beier U, Caiola N, *et al.* Patterns in species richness and endemism of European freshwater fish. Glob. Ecol. Biogeogr. 2007; 16: 65-75.

Sánchez-Montoya MM, Vidal-Abarca MR, Suárez ML. Comparing the sensitivity of diverse macroinvertebrate metrics to a multiple stressor gradient in Mediterranean streams and its influence on the assessment of ecological status. Ecol. Indic. 2010; 10: 896-904.

Singh KP, Basant A, Malik A, Jain G. Artificial neural network modeling of the river water quality--A case study. Ecol. Model. 2009; 220: 888-895.

Siroky DS. Navigating Random Forests and related advances in algorithmic modeling. Statist. Surv. 2009; 3: 147-163.

Smith KG, Darwall WRT, editors. The status and distribution of freshwater fish endemic to the mediterranean basin. Gland, Switzerland/Cambridge, UK.: IUCN -The World Conservation Union; 2006.

Strayer DL, Dudgeon D. Freshwater biodiversity conservation: recent progress and future challenges. J. N. Am. Benthol. Soc. 2010; 29: 344-358.

Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 2003; 43: 1947-1958.

Tan Y, Van Cauwenberghe A. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intell. 1999; 12: 21-35.

Tirelli T, Pessani D. Use of decision tree and artificial neural network approaches to model presence/absence of *Telestes muticellus* in piedmont (North-Western Italy). River Res. Appl. 2009; 25: 1001-1012.

Tirelli T, Pessani D. Importance of feature selection in decision-tree and artificial-neural-network ecological applications. Alburnus alburnus alborella: A practical example. Ecol. Inform. 2011; 6: 309-315.

Tirelli T, Pozzi L, Pessani D. Use of different approaches to model presence/absence of *Salmo marmoratus* in Piedmont (Northwestern Italy). Ecol. Inform. 2009; 4: 234-242.

Townsend C, Begon M, Harper J. Essentials of Ecology. Oxford: Wiley-Blackwell; 2008.

van Jaarsveld AS, Freitag S, Chown SL, Muller C, Koch S, Hull H, *et al.* Biodiversity assessment and conservation strategies. Science 1998; 279: 2106-2108.

Vezza P, Comoglio C, Rosso M, Viglione A. Low flows regionalization in North-Western Italy. Water Resour. Manag. 2010; 24: 4049-4074.

Vezza P, Parasiewicz P, Rosso M, Comoglio C. Defining minimum environmental flows at regional scale: application of mesoscale habitat models and catchments classification. River Res. Appl. 2012; 28: 675-792.

Vincenzi S, Zucchetta M, Franzoi P, Pellizzato M, Pranovi F, De Leo GA, *et al.* Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. Ecol. Model. 2011; 222: 1471-1478.

Wells B, Yu C, Koroukian S, Kattan M. Comparison of variable selection methods for the generation of parsimonious prediction models for use in clinical practice. In: Proceedings of the 33rd Annual Meeting of the Society for Medical Decision Making, Chicago, US; 2011.

Xu L, Zhang W-J. Comparison of different methods for variable selection. Analytica Chimica Acta 2001; 446: 475-481.

Chapter 5

Modelling factors affecting the presence/absence and density of *Luciobarbus guiraonis* (Júcar River Basin, Spain)

Esther Julia Olaya-Marín[1], Francisco Martínez-Capel[1], Rafael García-Bartual[2] and Paolo Vezza[1]

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

[2]Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Spain.

ABSTRACT

Luciobarbus guiraonis (Eastern Iberian barbel) is an endemic fish species to Spain, with the status of vulnerable species, threatened with extinction, mainly distributed in the Júcar River Basin District. Its study is important because there is scarce information about its biology and ecology. To improve the knowledge about the species distribution and habitat requirements, nonlinear modelling was carried out to predict the presence/absence and density of Eastern Iberian barbel. The models were created and validated against 155 sampling sites distributed throughout the Júcar River Basin District (Eastern Iberian Peninsula). In this case study, we used multilayer feed-forward artificial neural networks (ANN) to represent nonlinear relationships between *L. guiraonis* descriptors and various biological and habitat variables. The gradient descent algorithm was implemented to find the optimal model parameters and the importance of the ANN's input variables was determined by the partial derivatives method (PaD). The predictive power of the model was evaluated based on the Kappa statistic (*k*), the correctly classified instances (CCI), and the area under the curve (AUC) of a receiver operator characteristic (ROC) plots. According to the results, the presence/absence of *L. guiraonis* is well predicted by the ANN model (CCI= 87%, AUC= 0.85 and *k*= 0.66). The prediction of density was moderate (CCI = 62%, AUC=0.71 and *k*= 0.43). The most significant variables that described the presence/absence were: solar radiation (its highest contribution was observed between 2000 and 4200 WH/m^2), drainage area (with the strongest influence between 3000 and 5.000 km^2), and proportion of exotic fish species (with relevant contribution between 50 and 100%). In the density model, the most important variables were coefficient of variation of mean annual flows (relative importance of 50.5%) and proportion of exotic fish species (24.4%), but partial derivative method was unable to identify a positive or negative relationship between these variables and fish density. The models provide important information about the relation of *L. guiraonis* with biotic and habitat variables, this new knowledge could be used to support future studies and practical decisions about the management and conservation of this species in the Júcar River Basin District.

Keywords: Artificial Neural Networks; hydromorphology; species distribution model; Mediterranean rivers; cyprinids; fish habitat; fish density.

5.1 INTRODUCTION

Models are useful to understand the effects of environmental variables interactions on species distribution and to assess the alteration of ecological patterns as a response to environmental changes (Jopp *et al.*, 2011). Ecological modelling have become an important tool to learn about the implications of stressors like climate change, hydrological regime alteration, water pollution, and invasive species introduction on freshwater ecosystems (Drew *et al.*, 2011). Ecological models can be integrated in support decision making systems for ecological restoration, reserve design and conservation planning, impact assessment and resource management (Drew *et al.*, 2011; Franklin, 2010; Guisan and Thuiller, 2005; Olaya-Marín *et al.*, 2012; Pearce and Ferrier, 2000). This models are needed in Mediterranean rivers (Olaya-Marín *et al.*, 2012; Vezza *et al.*, 2012), because 56% of endemic fish are threatened with extinction (Smith and Darwall, 2006). From a conservation point of view, models of species density and presence/absence help to search for better conservation and river restoration politics (Costa *et al.*, 2012; Muñoz-Mas *et al.*, 2012; Zarkami *et al.*, 2012).

Machine learning (ML) techniques have been seen in recent years as a promising discipline to advance the current knowledge of ecological processes and patterns (Drew *et al.*, 2011; Jopp *et al.*, 2011; Leclere *et al.*, 2011; Olden *et al.*, 2008). ML has been developed from artificial intelligence and has been applied in several disciplines of environmental sciences (Hsieh, 2009), due to their ability to model nonlinear processes (Hsieh, 2009; Olden *et al.*, 2008). This ML feature allows us to derive better predictions and improve the effectiveness of decision making in environmental management (Evans and Cushman, 2009).

Currently, ML is a cornerstone and one of the most active research areas in the field of artificial intelligence (Jopp *et al.*, 2011). Artificial neural networks (ANN) are one of the most effective ML techniques capturing nonlinearities in ecological problems (Feio and Poquet, 2011; Franklin, 2010; Lek *et al.*, 2005), the use of ANN in freshwater studies demonstrates this affirmation. Baran *et al.* (1996)

predicted density and biomass of *Salmo trutta* in some rivers of the central Pyrenean mountains, Mastrorillo *et al.* (1997) modelled presence/absence of three fish species in south-western France, Brosse *et al.* (1999) found a good performance of ANN to predict fish spatial occupancy and abundance in a mesotrophic reservoir, Joy and Death (2004) simulated fish and decapod presence in Wellington Region North Island New Zealand. Fish presence/absence is well classified by ANN both in temperate zones (Tirelli and Pessani, 2011; Tirelli *et al.*, 2009) and tropical ones (Hauser-Davis *et al.*, 2010), even modelling with unbalanced datasets (in terms of prevalence). Olaya-Marin *et al.* (2012) found satisfactory results modelling native fish richness with ANN, their approach provided an evaluation of the effects of hydromorphological changes and river restoration actions (weir removal) in three Spanish Mediterranean rivers.

This study focused on Eastern Iberian barbel (*Luciobarbus guiraonis*), because it is a Mediterranean endemic freshwater fish mainly distributed in the Júcar River Basin, in Spain (Doadrio, 2001; Jiménez *et al.*, 2002), and their ecology and biology is poorly known (Doadrio, 2001); Moreover, this species is facing a high risk of extinction and its population is estimated to decline by 30% in the next ten years (IUCN, 2012). The ichthyofauna of the Iberian Peninsula is one of the most endemic in the world (Doadrio and Aldeguer, 2007); therefore, the study of aquatic ecosystems in this area should be supported by reliable models to predict future ecological changes, in order to understand potential alterations and avoid their occurrence through the implementation of restoration actions (Clark *et al.*, 2001; Drew *et al.*, 2011; Olaya-Marín *et al.*, 2012).

In this paper, we evaluate the ability of ANN to identify local stressing factors affecting density and presence/absence of *L. guiraonis* in the Júcar River Basin District at the basin scale. The aims of the research are (i) to identify relevant environmental variables and model density and presence/absence of *L. guiraonis* and (ii) to assess the importance of each predictive environmental variable in the estimation of density and presence/absence.

5.2 MATERIALS AND METHODS

5.2.1 STUDY AREA AND DATA COLLECTION

The research was conducted in the main streams of the Júcar, Cabriel and Turia rivers (Eastern Iberian Peninsula). This area is characterized by torrential storms in humid seasons, severe droughts in dry seasons and a hydrological response controlled by the rainfall regime, which is typical in Mediterranean environments (Granado-Lorencio, 1996; Vila-Gispert *et al.*, 2005). Temporal irregularities of rainfall cause a particular variability pattern expressed by interannual differences in discharge (Gasith and Resh, 1999; Granado-Lorencio, 1996). The natural hydrological regime in these rivers has been altered by dams and weirs, used for hydropower generation and water consumption for agricultural, industrial and domestic activities. The most altered habitats are located in the middle and lower parts of the watersheds, in which there is a poor development of riparian vegetation caused by agricultural pressures and wastewater discharges (Martínez-Capel *et al.*, 2008), this situation is mainly evidenced in stretches located from Alarcón's dam to the mouth in Júcar river.

Mediterranean fish communities are known by their low species richness, high endemicity (Ferreira *et al.*, 2007) and the predominance of Cyprinids, which possess high specific diversity, are exclusive of epicontinental water bodies, and have typical morphofunctional and physiological adaptations to fluctuating environments (Granado-Lorencio, 1996, 2000). The current knowledge about Mediterranean freshwater fish inhabiting the Iberian Peninsula is scarce (Aparicio *et al.*, 2011; Ferreira *et al.*, 2007; Maceda-Veiga and De Sostoa, 2011). Whereas other studies about habitat requirements at microhabitat and mesohabitat scale are available in the Iberian Peninsula (e.g. Costa *et al.*, 2012; Martínez-Capel *et al.*, 2009), the studies at the basin scale are very scarce. *L. guiraonis* is a cyprinid species and it has been poorly studied; consequently, there are little information about their biology and ecology, inferred from similar species like *Luciobarbus graellsii* and *Luciobarbus bocagei* (Doadrio, 2001). But these extrapolations could conduct erroneous conclusions about the ecology of *L. guiraonis* and the design of inadequate restoration measurements and management actions (Aparicio *et al.*, 2011).

5.2.2 RESPONSE AND PREDICTIVE VARIABLES

The sampling dataset is constituted by 145 sites along the main streams of the considered Mediterranean rivers (Fig. 1). Density and presence/absence of *L. guiraonis* were calculated based on single-pass electrofishing in spring, summer and fall seasons from 2004 to 2010. Fish density and presence/absence were the dependent variables for the models in function of biological and habitat variables. Species density was expressed as the number of fish caught per m^2.

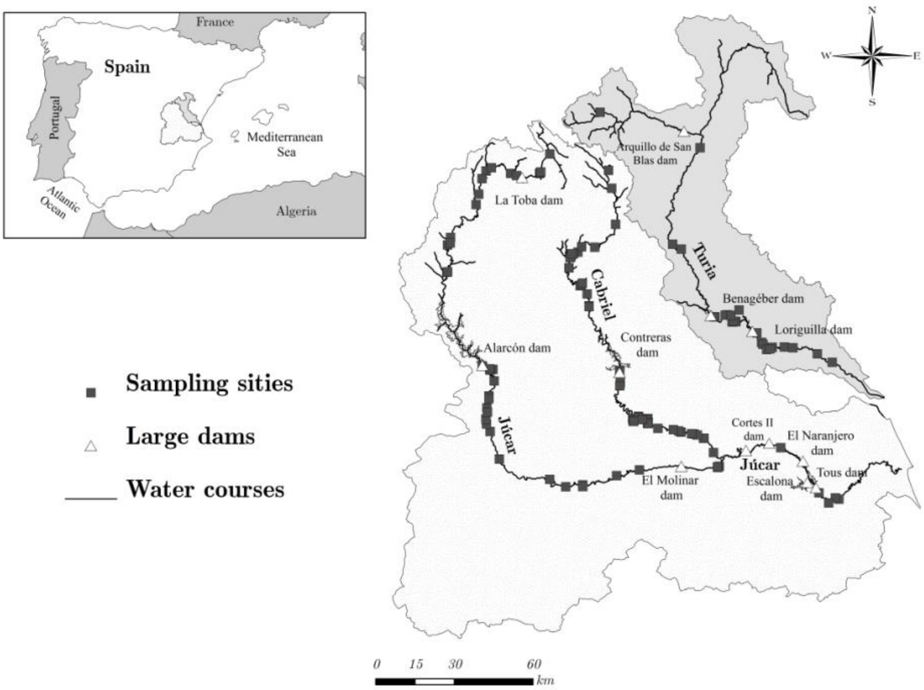

Fig.1. Study area showing the distribution of the 145 sampling sites in the three rivers (Jucar, Cabriel and Turia rivers).

Twenty seven environmental variables were used as potential predictors in the ANN models (Table 1), these variables were selected taking into account their ecological importance for fish life cycle (Bernardo *et al.,* 2003; Costa *et al.,* 2012; Granado-Lorencio, 1996; Jackson *et al.,* 2001; Oberdorff *et al.,* 1995). The potential predictors were obtained from three sources (Table 1): in situ (field work), GIS analysis and official monitoring networks (MN) of stream flow and biological variables. Geographical variables (i.e. altitude, distance from the source, channel length without artificial barriers, and others) were computed in ArcGIS 9.3.1 with a 5-meter resolution digital elevation model supplied by the National Geographical Institute of Spain.

Altitude, water temperature, and the natural slope in the channel play an important role in fish communities' distribution along the rivers, since they influence flow velocity, water oxygenation and the magnitude of sediments transported by the stream (De Sostoa, 2002; Jackson *et al.,* 2001), factors that affect the development of different habitats for fish life (Bernardo *et al.,* 2003; Costa *et al.,* 2012). Channel length without artificial barriers, number of tributaries between artificial barriers and drainage area between artificial barriers were included in the research because dams and weirs are physical obstacles limiting fish migration along the river (García de Jalón and González del Tánago, 2007; García de Jalón *et al.*, 2007); moreover, longitudinal connectivity restoration is critical to re-establish the natural dynamics of freshwater ecosystems (Lake *et al.,* 2007). Other geographical and hydrological variables like distance from the source, drainage area, potential insolation, magnitude and variability of river discharge are key factors for Mediterranean ichthyofauna conservation (Filipe *et al.,* 2010; Granado-Lorencio, 2000; Hermoso and Clavero, 2011).

It is well known that water quality affects the distribution and composition of fish communities (Jackson *et al.,* 2001; Schlosser, 1991). We considered water quality indices (ICGp, IBMWP and QBR) because they give an integrative estimation based on different source of environmental quality and helped us to reduce dimensionality. The Iberian Biomonitoring Working Party index (IBMWP) is a modification of the Biological Monitoring Working Party score system (1978), adapted by Tercedor and Sánchez-Ortega (1996; 1988) to the Iberian Peninsula.

This index assesses biological quality in water bodies based on macroinvertebrates, their values are obtained by the sum of the partial scores assigned to each family of macroinvertebrates present in each stretch. The total IBMWP score goes from 0 to more than100 (it is possible to find IBMWP values higher than 300 in Iberian rivers). The general physico-chemical water quality index (ICGp) is a variation of the general quality index (CHJ, 2008; Martínez-Muro, 2003), originally developed by Provencher and Lamontagne (1977). The ICGp index results from the combination of 11 parameters, its values are limited in the range of 0 (heavily polluted water) to 100 (very good quality).

The interactions among species influence fish distribution (Broennimann *et al.,* 2012; Fitzpatrick *et al.,* 2007), but this kind of variables are commonly neglected when species distribution models are built (Davis *et al.,* 1998; Fitzpatrick *et al.,* 2007). However, we have included in our analysis some variables related to the fish communities and species interactions, like the proportion of exotic fish species (%), total density of invasive fish, and fish predators' density, since exotic species are one of the main causes of threat for endemic ichthyofauna in Mediterranean rivers (Doadrio, 2001b; Granado-Lorencio, 1996; Smith and Darwall, 2006).

At ungauged locations, the mean monthly flow was calculated through a linear interpolation based on the relationship among flow in natural conditions and the accumulated drainage area between gauged sites (Caissie, 2006a; Caissie and El-Jabi, 1995; Leopold and Maddock, 1953; Leopold *et al.,* 1964). River discharge and its pattern of variability define the lifecycle traits of Mediterranean fish species (Ferreira *et al.,* 2007) therefore, different variables regarding the magnitude and variability of river flows (at monthly scale) during the previous years and in the spawning season before the fish sampling were considered. The riparian habitat quality index (QBR, Munné *et al.,* 2003) was introduced in the study to assess the morphological conditions of the sampling sites. QBR is the integration of four components and synthesizes different qualitative features describing the conservation status of the riparian area: total vegetation cover, vegetation cover structure, vegetation cover quality, and river channel alterations. Each feature can be evaluated from 0 to 25 and the total valuation can be in the range of 0 to 100. QBR is accepted as a well approximation of riparian quality in different regions, including Mediterranean environments (Aguilella *et al.,* 2005; Garófano-Gómez *et al.,* 2011). Finally, fishing year and river name were used as potential predictive variables in order to know if these variables have a relevant weight predicting density and presence/absence of *L. guiraonis*.

Table 1. Potential environmental variables selected to build the presence/absence and density predictive models.

Variable	Code	Method	Range	Mean
Channel length without artificial barriers (km)	CWB	GIS	0.8-79.0	43.8
Number of tributaries between artificial barriers	TAB	GIS	0.0-54	28.7
Altitude (m a.s.l)	ALT	GIS	28-1286	553.8
Drainage area (km^2)	DRA	GIS	95-18296	4189.0
Drainage area between artificial barriers	DAB	GIS	3.0-4624	866.0
Distance from the source (km)	DHS	GIS	20.53-383.7	168.0
Natural slope of the channel (%)	NSL	GIS	0.0-6.8	0.44
Solar radiation (WH/m^2)	SOR	GIS	1153-6298	3915.6
Water temperature (ºC)	WAT	MN	5.76-19.9	13.6
Mean Annual flow rate (m^3/s)	FMA	MN	0.03- 12.22	5.1
Mean monthly flow (Two year before sampling) (m^3/s)	MMF		0.103- 13.62	5.13
Inter-annual mean flow (5 years before sampling)	FIA	MN	0.11- 12.36	6.02
Coefficient of variation of mean monthly flows (5 years before sampling)	FIM	MN	0.23- 1.09	0.65
Coefficient of variation of mean annual flows (5 years before sampling)	FCV	MN	0.15-0.91	0.36
Inter-monthly flow variation of the mean monthly flows (5 years before sampling)	FVM	MN	0.36-3.37	0.83
Maximum monthly flow (April to June before sampling) (m^3/s)	MaxMF	MN	0.015- 26.38	7.82
Mean monthly flow (April to June before sampling) (m^3/s)	MeanMF	MN	0.01- 21.0	6.3
Minimum monthly flow (April to June before sampling) (m^3/s)	MinMF	MN	0.0015-7.87	2.73
Mean monthly flow (of the two months with the lowest monthly flow for the year before sampling) (m^3/s)	MeanLMF	MN	0.004- 8.2	2.9
Sampling year	FIY	n/a	2004-2010	n/a
River name	RN	n/a	1-3	n/a
Proportion of exotic fish species (%)	PEF	*In situ*	0.0-100	37.6
Total density of invasive fish ($Fish/m^2$)	DIF	*In situ*	0.0-0.008	0.002
Fish predator density ($Fish/m^2$)	FPD	*In situ*	0.0- 0.0035	0.0005
Index of Riparian Habitat Quality	QBR	MN	10-100	73.28
Iberian Biomonitoring Working Party	IBMWP	MN	61-260	124.3
Physicochemical Index of water quality	ICGp	MN	67.57-87	80.5

5.2.3 ARTIFICIAL NEURAL NETWORKS MODELLING

Artificial neural networks are mathematical models inspired in nervous system structure, being its fundamental building block is the neuron (Lek *et al.*, 2005; Olden *et al.*, 2008). ANN are valuable in ecological studies because they are flexible, robust, and generalizable (Alpaydın, 2010; Lek *et al.*, 2005; Olden *et al.*, 2008). The most used ANN in ecology is the multilayer perceptron (MLP) (Özesmi *et al.*, 2006), which can be successfully applied in pattern recognition problems, forecasting, signal processing and modelling of complex nonlinear systems (Goethals *et al.*, 2007; Lek *et al.*, 2005). In the learning phase, MLP weights are updated to reduce the differences between observed and predicted outputs; this process ends when the stopping criterion is reached (e.g. *early stopping*: which implies stopping the training phase when validation error increases in a specified number of iterations). Given the advantages of MLP, in this study we used it to predict presence/absence and density of *L. guiraonis*. Presence/absence was considered as a binary variable, where presence was denoted by 1 and absence was represented by 0. On the other hand, density was classified according to the number of fish per square meter; class 1 means a density of zero, class 2 comprises densities between 0.001 and 0.019, and class 3 comprises densities in the range from 0.020 to 0.066 ind/m^2 (Table 2).

Table 2. Range and total number of data classes used to build the presence/absence and density predictive models.

Model	Class	Range	Number of data
Presence/Absence	Presence	1-1	102
	Absence	0-0	43
Density	Class 1	0-0	43
	Class 2	0.001 -0.019	52
	Class 3	0.020 -0.066	50

The generalization capacity of ANN can be restricted by the distribution pattern and magnitudes of the original data. Data pre-processing is highly recommended before building the ANN models (Goethals *et al.*, 2007). Accordingly, input variables have been previously transformed to a similar order of magnitude. In this case, environmental variables involved in the problem were proportionally scaled between 0 and 1 in the range of their maximum and minimum values (e.g. Olden and Jackson, 2002b; Park *et al.,* 2008; Qin *et al.,* 2010; Tirelli and Pessani, 2009). Another aspect playing a crucial role in the overall performance of the resulting network is the input variable selection process. A two-step methodology has been used herein. Firstly, an exploratory analysis to identify collinearity among the potential predictors was carried out by hierarchical cluster analysis using squared Spearman correlations (ρ^2) as similarity measure (Fig. 2). In the case of highly correlated variables ($\rho^2 > 0.8$), only the one with higher ecological interpretability was chosen (Olaya-Marín *et al.,* 2012). Secondly, a forward stepwise method was conducted to eliminate irrelevant inputs and thus, reducing network architecture complexity (Gevrey *et al.,* 2003).

Several MLP models were built and tested, in order to establish (by systematic trial and error) the optimal number of neurons in the hidden layer and the proper transfer function in the hidden and output layers. Commonly, transfer functions are nonlinear; they transform the weighted sum of inputs into an output signal (Isa *et al.,* 2010; Zhang *et al.,* 1998) and it is typical the use of the same transfer function in hidden and output layers (Goethals *et al.,* 2007; Lek *et al.,* 2005). MLP results are very sensitive to the implemented transfer functions in their layers (Isa *et al.,* 2010; Piekniewski and Rybicki, 2004). Generally, the selection of a transfer function is based on the best performance by trial and error (Isa *et al.,* 2010), comparing different transfer functions in the hidden and output layers. In this work, two transfer function combinations (hidden layer/output layer) were tested: hyperbolic tangent/linear, and logistic/linear; the combination offering the best performance was selected (Isa *et al.,* 2010; Olaya-Marín *et al.,* 2012).

The ANN models were designed with a single hidden layer and the number of neurons optimized by trial and error. Bishop (1996) have shown that a single hidden layer is sufficient for statistical applications with reasonable computation requirements. Moreover, the use of a single hidden layer is comparable to the results using multiple hidden layers (Kurková, 1992). The dataset was randomly divided into three sections (training: 60%, validating: 20% and testing sets: 20%), these percentages are frequently used in literature (Qin *et al.,* 2010; Ryan *et al.,* 2004). The gradient descent with momentum and adaptive learning rate backpropagation algorithm (*traingdx* function in Matlab) was the optimization method used to train the networks (Demuth *et al.,* 2010) with a momentum constant of 0.9 and a learning rate of 0.01. Three efficiency indices have been selected to evaluate model's predictive capacity: Correctly classified instances (CCI) (Buckland and Elston, 1993; Fielding and Bell, 1997), Cohen's Kappa (Cohen, 1960) calculated upon the confusion matrix (Table 3) and the area under the curve (AUC) of the receiver-operator characteristic (ROC) plots (Hanley and McNeil, 1982).

Table. 3. Error matrix used to calculate the percentage of correctly classified instances (CCI) and Cohen's Kappa.

	Observed	
Predicted	Presence	Absence
Presence	a (true positive)	b (false positive)
Absence	c (false negative)	d (true negative)

The percentage of correctly classified instances (CCI) was calculated as follows:

$$CCI = \left(\frac{a+d}{n}\right) \times 100 \quad (1)$$

Cohen's kappa measures the proportion of correctly classified points after accounting for the probability of chance agreement (Drew *et al.,* 2011). Kappa ranges from 0 to 1. According to previous researches (Drew *et al.,* 2011; Koch *et al.,* 1977; Manel *et al.,* 2001) the index can be valued as poor (0.00 to 0.39), moderate (0.40 – 0.59), substantial (0.60 to 0.79) or excellent (0.80 to 1). Cohen's kappa was calculated by the following equation:

$$Kappa = \frac{\left(\frac{a+d}{n}\right) - \left(\frac{(a+b)(a+c)+(c+d)(b+d)}{n^2}\right)}{1 - \left(\frac{(a+b)(a+c)+(c+d)(b+d)}{n^2}\right)} \quad (2)$$

AUC is calculated as the area under the ROC curve (Franklin, 2010) and it is applicable only to binary variables (e.g., presence/absence). AUC is interpreted as the probability of correctly classifying a pair of subjects randomly selected, one from the presence group and the other from the absence group (Fielding and Bell, 1997; Franklin, 2010). AUC ranges from 0.50 to 1, a value between 0.50 and 0.70 indicates a low discrimination (poor model performance), from 0.70 to 0.90 moderate discrimination and higher than 0.90 high discrimination (Manel *et al.*, 2001; Swets, 1988).

The selection of a parsimonious model is important to find a robust model. We took into account the number of parameters using the Akaike's Information Criterion (AIC); this performance metric considers the fitting error and the number of variables used to reach that error, and is useful to assess the relationship between fitting and neural network size. AIC is calculated as follows:

$$AIC = n\ln(RMSE) + 2k \tag{3}$$

Where $\ln$ is the natural logarithm, k is the number of network weights, calculated as follows: k = input neurons × hidden neurons + hidden neurons × output neurons + bias parameters (see Brosse and Lek, 2000, for details), n is the sampling size and RMSE is the root mean squared error. A smaller AIC means a better performance in relation to the number of parameters used by the model. So, choosing the model with the smallest AIC implies to select the simplest one with less inputs and hidden neurons. This criterion is valuable due to the models with fewer variables are much easier to interpret, have a lower level of prediction uncertainty (Jorgensen and Fath, 2011), and optimal ANN architecture is the simplest one that adequately captures the relationships in the training data (D'Heygere *et al.*, 2006).

Finally, to evaluate the importance of input variables in each model, the partial derivatives method (PaD) was implemented (Dimopoulos *et al.*, 1995). PaD method can be used to analyse the output changes as a response of small variations in each input variable, and estimate input relative importance to predict presence/absence and density.

5.3 RESULTS

The correlation matrix of variable indicated that altitude (ALT) and drainage area (DRA) were strongly correlated (Fig. 2). Following literature, DRA has a higher ecological importance for fish and is broadly used to explain variations among aquatic communities (Ibarra *et al.*, 2003; Olaya-Marín *et al.*, 2012); thus, ALT was removed as potential predictive variable. Minimum monthly flow (MinMF, April-June) was highly correlated (ρ^2 = 0.97) with mean monthly flow of the lowest 2 flows (MeanLMF); MeanLMF was omitted as potential predictive variable because MinMF acts as a critical threshold for fish spawning. Maximum monthly flow (MaxMF) has a correlation of 0.98 with mean monthly flow (MeanMF); because MaxMF is important to *L. guiraonis* spawning, which occurs from April to June (Doadrio, 2001), we excluded MeanMF from the dataset. According to Fig. 2, the number of tributaries between artificial barriers (TAB) had a strong correlation with the channel length without artificial barriers (CWB); however, we preserve both of them for input selection through forward stepwise method, given their importance for Mediterranean fish life (Granado-Lorencio, 2000; Olaya-Marín *et al.*, 2012).

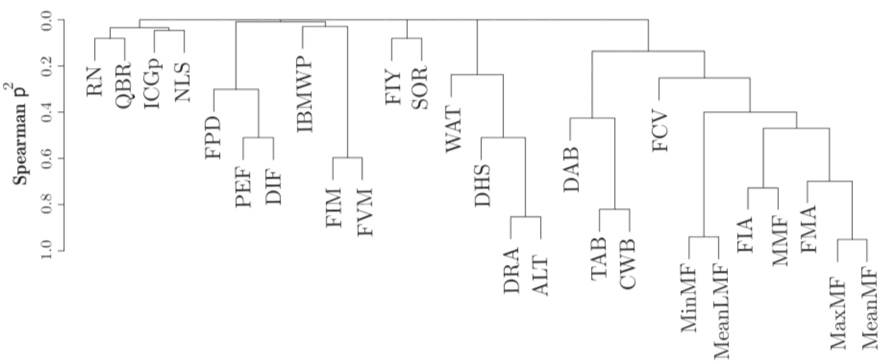

Fig. 2. Hierarchical clustering using squared Spearman correlation (ρ^2) of environmental variables. Nomenclature is shown in Table 1.

Hyperbolic tangent in the hidden layer and linear in output layer were the best transfer functions found to predict presence/absence of *L. Guiraonis*. Fig. 3a shows that Cohen's Kappa increases at a high rate from 1 to 8 predictive variables, the rate of improvement is dramatically reduced from 8 to 19 variables, and using more than 19 variables there is another significant increment (maximum k=0.85). The density model (Fig. 3b) reached the best performance using logistic and linear transfer functions; the best performance with a small number of inputs was reached by the model with 5 input variables, and using more than 14 variables we can see a better performance, but we selected the model with 5 predictors because a model with a reduced number of input variables is more interpretable and applicable (Drew *et al.*, 2011).

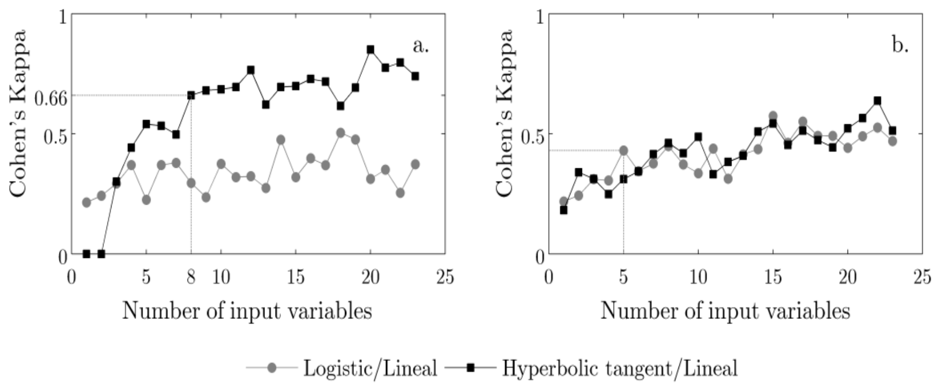

Fig. 3. Influence of the number of input variables and transfer functions (Hidden layer/output layer) in models' performance. a) presence/absence of L. guiraonis. b) Density of L. guiraonis.

To predict presence/absence of *L. guiraonis*, the best neural network architecture found had three layers (i.e. 8→6→2) with eight nodes in the input layer, six neurons in the hidden layer, and two neurons in the output one (Fig. 4a). The model with 12-inputs (12→9→2) has a kappa of 0.77 (Fig. 3 and Table 4), but it was not selected because it has almost twice the number of parameters of the 8-inputs model. This involves a test-AIC of 243.88 in the 12-inputs model and a test-AIC of 113.4 in the 8-inputs model (Table 5). Moreover, both models have a kappa coefficient classified as substantial.

The eight predictive variables identified in the selected model were: Drainage area (DRA), proportion of exotic fish species (PEF), solar radiation (SOR), mean annual flow rate (FMA), number of tributaries between artificial barriers (TAB), coefficient of variation of mean monthly flow (FIM) and the Iberian Biomonitoring

Working Party (IBMWP). This network possesses a high percentage of CCI (87%), a substantial kappa coefficient (0.66) and a good performance evidenced by a value of 0.85 in the area under the ROC curve, which indicate that this model discriminates well (Gabriels *et al.*, 2007; Manel *et al.*, 2001).

Table 4. Predictive results of ANN models (CCI= Percentage of correctly classified instances; Cohen's kappa; AUC= Area Under the Curve)

Model	ANN-structure	CCI	*Cohen's kappa*	AUC
Presence/absence	8→6→2	87%	0.66	0.85
	12→9→2	90%	0.77	0.93
Density	5→4→3	62%	0.43	0.71
	15→11→3	72%	0.57	0.83

We found a model with five inputs as the best parsimonious tool to interpret density of *L. guiraonis* (Table 5). These inputs are the coefficient of variation of mean annual flows (FCV), proportion of exotic fish species (PEF), minimum monthly flow from April to June before sampling (MinMF), sampling year (FIY) and IBMWP. The architecture of this model is composed by five nodes in the input layer, four neurons in the hidden and three nodes in the output one (i.e., 5→4→3), the model classifies density in three classes in the output layer (Fig. 4b). This model presents a lower performance than presence/absence model (Table 4); with CCI of 62%, a moderate kappa coefficient (0.43) and an AUC of 0.71. These values indicate a moderate efficiency of the model (e.g. Tirelli *et al.*, 2009).

Models with more input variables also have a moderate efficiency. For example, density model with 15 inputs (15→11→3) presents a moderate kappa coefficient (0.57), a moderate AUC (0.83) and a CCI of 72%. However, this model has 212 parameters in contrast to the 39 parameters of the 5-inputs density model. Table 5 shows a better performance concerning AIC in the 5-inputs density model than 15-inputs density model.

Table 5. Akaike Information Criterion (AIC) calculated in presence/absence and density models (RMSE= Root mean squared error, n = sampling size, k = neural network parameters)

Model		Phase	RMSE	*n*	*K*	AIC
Presence/Absence	ANN (8→6→2)	Training	0.327	87	68	38.76
		Validation	0.310	29		102.08
		Test	0.466	29		113.84
	ANN (12→9→2)	Training	0.224	87	137	143.8
		Validation	0.433	29		249.73
		Test	0.354	29		243.88
Density	ANN (5→4→3)	Training	0.422	87	39	2.92
		Validation	0.452	29		54.95
		Test	0.439	29		54.17
	ANN (15→11→3)	Training	0.375	87	212	338.56
		Validation	0.419	29		398.75
		Test	0.400	29		397.45

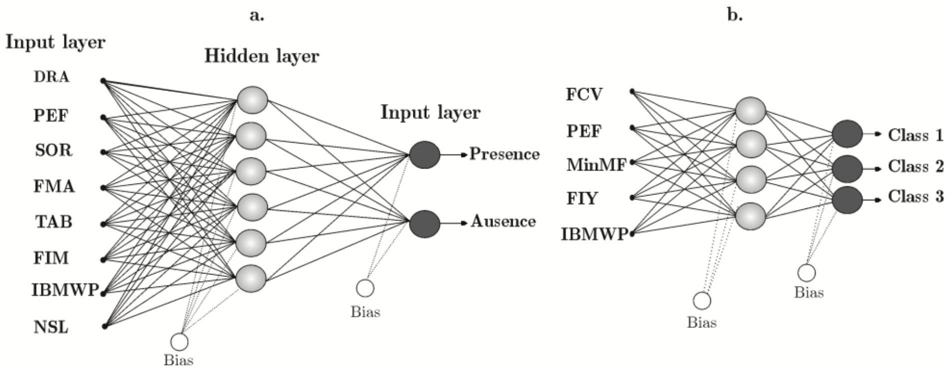

Fig. 4. Schematic description of the ANN models finally selected. a) presence/absence of L. guiraonis. b) Density of L. guiraonis. Variables' nomenclature is presented in Table 1.

The implementation of the partial derivatives algorithm revealed that the most influential variables to predict presence/absence of *L. guiraonis* were solar radiation (SOR) with a relative importance of 27.8%, drainage area (DRA) with 24.53%, and proportion of exotic fish species (PEF) with 13.60% (Fig. 5a). Partial derivatives of each of these variables were plotted against the corresponding input values (e.g. Brosse *et al.,* 2003). Positive values in y axes (Fig. 6) indicate a positive relationship between the input and the output variable, on the contrary, negative values express an inverse relationship (Gevrey *et al.*, 2003; Olaya-Marín *et al.*, 2012).

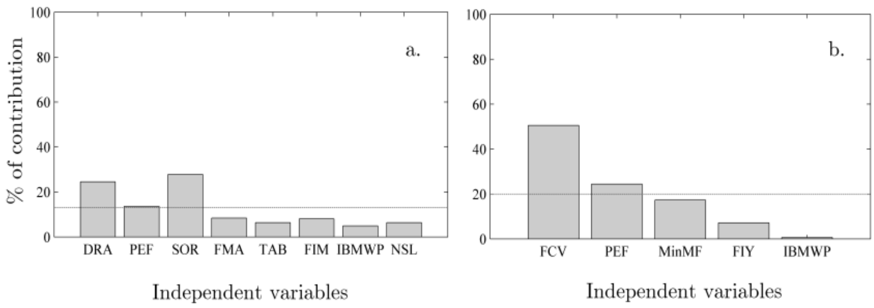

Fig. 5. Contribution of each independent variable predicting presence/absence (left) and density (right), based on the partial derivatives method (PaD). Dotted lines represents the level of significance (13 and 20% respectively) according to Brosse et al, (2003).

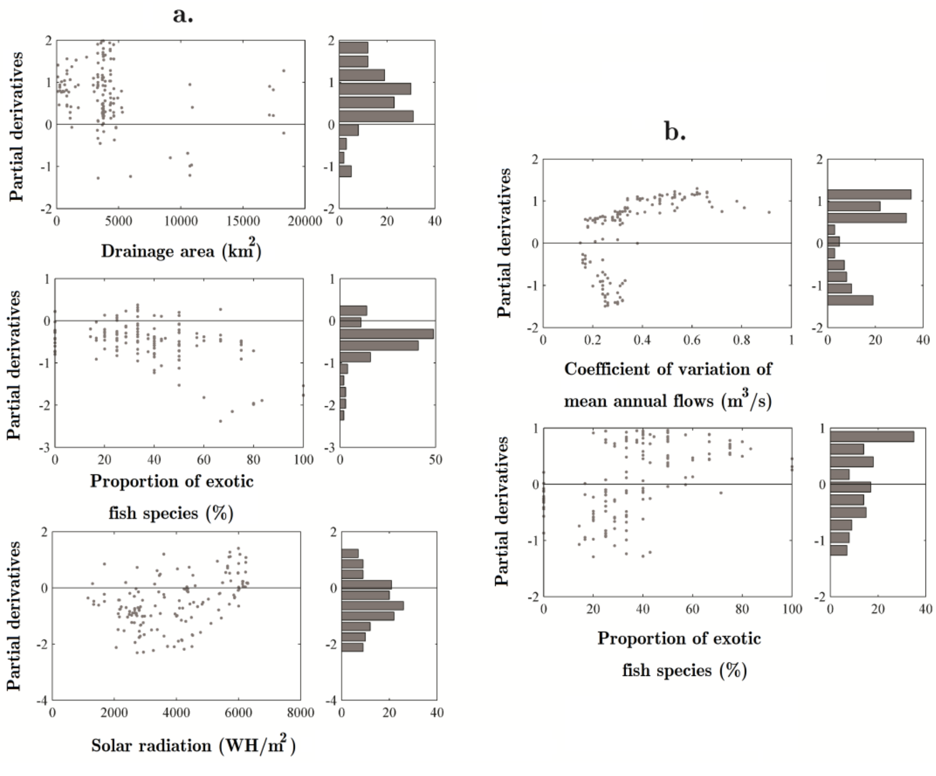

Fig. 6. Output partial derivatives related to the most significant predictive variables, as a function of each environmental variable. Left a) Presence of L. guiraonis. Right b) Density of L. guiraonis.

The analysis of Fig. 6a leads to the following remarks:

- Most positive partial derivatives with respect to drainage area (DRA) occurs in the range between 3000 to 5000 km^2. This means that the increase of DRA would conduct to an increment of *L. guiraonis* presence, but this relation is weaker in DRA values higher than 5000 km^2.

- When the proportion of exotic fish species (PEF) reaches the range of 50 to 100% we found the highest negative partial derivatives. So, for values of PEF higher than 50 we can have a higher importance of PEF in the reduction of *L. guiraonis* presence.

- The negative partial derivatives in relation to solar radiation (SOR) show that the increase of the SOR contributes to the reduction of probability of *L. guiraonis* presence, and the highest contributions are found in the range of 2000 to 4200 WH/m^2.

The most important variables in the density model were the coefficient of variation of the mean annual flow (FCV) with a relative importance of 50.5% (Fig. 5b), and the proportion of exotic fish species (PEF) with 24.40%. Given the moderate performance of this model, PaD method is unable to support an ecological interpretation of the predictive variables. Partial derivatives with respect to FCV and PEF are positive and negative at the same time, without a precise tendency; therefore, it is not possible to provide a robust mathematical conclusion about the action of FCV and PEF on *L. guiraonis* density (Fig. 6b).

5.4 DISCUSSION AND CONCLUSION

The developed model to classify the presence/absence of the studied species is efficient, and their predictive variables have ecological relevance; moreover, the relationships revealed by the PaD method have ecological meaning. These features increase the interest of the proposed model, because it can be used to study the relationships between the environmental variables and *L guiraonis*, which have a direct impact to improve the understanding of the fluvial Mediterranean ecosystem behaviour.

Regarding the selected density model, it is important to highlight that the selected inputs are relevant for the studied species; even if the model had a moderate performance; it is not possible to establish a clear relationship between input variables and species density from PaD analysis (Fig. 6b). It could be caused by an interaction between FCV and PEF or another variable (e.g. Gevrey *et al.,* 2003) which probably imply the difficulty to identify a single input perturbation effect on species density (Gevrey *et al.,* 2006). For this reason, it is recommended employ in future analysis PaD2 (see Gevrey *et al.*, 2006, for details) to study the contribution of all possible pair-wise combinations of input variables.

From the partial derivatives method, the predictive variables with the strongest contribution to predict presence/absence were drainage area (DRA), solar radiation (SOR) and proportion of exotic fish species (PEF). Based on our results, *L. guiraonis* have a positive relation with drainage area and the species prefers the stretches located at the middle parts of the watersheds (Fig 6a). Gortázar *et al.* (2007) found a low presence of *L. guiraonis* in the upper parts of Cabriel River and a large population in the middle, which is a zone with a proper temperature range for cyprinids development. Kottelat and Freyhof (2007) affirmed that the presence of *L. guiraonis* is more frequent in middle and lower river reaches, and related its decline with water abstraction and habitat modification; these alterations are present in the lower parts of the Júcar, Cabriel and Turia rivers. Martínez-Capel *et al.* (2008) discussed that the population of the studied species have declined in the Júcar River through the years, due to the large proportion of lentic habitats (produced by frequent weirs) and the high proportion of fine-textured soils in the channel bed, which affects the eggs survival. The same author revealed that the lack of recruitment can be the main cause of *L. guiraonis* declining, which is noticeable in the lower reaches of the Júcar and Turia rivers (Estrela *et al.,* 2004).

Concerning the solar radiation (SOR), the sensitivity analysis indicated a negative relationship between SOR and the presence of *L. guiraonis* (Fig. 6b). Several authors have found that solar radiation is a critical environmental factor governing temperature change in fluvial systems (Brown and Krygier, 1970; Isaak *et al.,* 2012; Webb *et al.,* 2008). Water temperature is a key variable for fish survival because affects their physiology and behaviour (Caissie, 2006b; Hrachowitz *et al.,* 2010). This environmental variable directly regulates dissolved oxygen concentration in the water, affecting spawning time, growing rates, and spatial-temporal distribution of species (Baron *et al.,* 2002; Jackson *et al.,* 2001; Magnuson *et al.,* 1979; Prchalová *et al.,* 2006). Currently, there is a deficiency of knowledge about the effect of water temperature on *L. guiraonis* and more efforts are needed to investigate this relationship because climate change studies have indicated that Mediterranean rivers would experience an increase of droughts intensity and frequency (Bonada *et al.,* 2007; Hermoso and Clavero, 2011; Mas-Martí *et al.,* 2010; Sabater and Tockner, 2010), which could severely affect the establishment and survival of this species in the future.

The negative relationship between the proportion of exotic fish species (PEF) and the presence of *L. guiraonis* (Fig. 6a) is supported by the findings of Doadrio (2001), who described a reduction of this species due to the introduction of exotic species. In the studied river basins, *L. guiraonis* cohabits with the following exotic species: Pumpkinseed (*Lepomis gibbosus*), Largemouth bass (*Micropterus salmoides*), Northern pike (*Esox lucius*), the pikeperch (*Sander lucioperca*), Pyrenean gudgeon (*Gobio lozanoi),* Bleak *(Alburnus alburnus),* Common carp (*Cyprinus carpio),* Iberian straight-mouth nase *(Pseudochondrostoma polylepis),* Gudgeon *(Gobio gobio)* and Rainbow trout *(Oncorhynchus mykiss).* Exotic species have a variety of adverse effects on native fauna, such as predation, competition, hybridization, disease vector and habitat alteration (Almeida and Grossman, 2012; Granado-Lorencio, 2000). In the Júcar, Cabriel and Turia rivers, these invasions have been favoured by the construction of dams and weirs, which have created a suitable habitat for the establishment of exotic species (Olaya-Marín *et al.,* 2012), changing from a lotic system to a lentic one, where some exotic species find the suitable habitat for spawning (e.g. the bass and the pumpkinseed). Moreover, these hydraulic structures have segregated *L. guiraonis* to isolated habitats and interrupted upstream migration in the spawning season. Generally speaking, many exotic species cannot be adapted to the natural Mediterranean hydrological regime, characterized by water stress in summer and a torrential regime in fall (Doadrio and Aldeguer, 2007); hence, they are present in the studied area because the flow regulation and artificial obstacles have benefited the establishment of these species (Corbacho and Sánchez, 2001; Vila-Gispert *et al.*, 2005).

The predictive variables of the density model (FCV, PEF, MinMF, FIY and IBMWP) have a reasonable influence on Mediterranean fish species from an ecological point of view (Bernardo *et al.,* 2003; Doadrio, 2001; Granado-Lorencio, 2000; Olaya-Marín *et al.*, 2012). Previous studies have demonstrated that the coefficient of variation of the mean annual flow (FCV) is important for fish distribution in Mediterranean rivers, because native fish lifecycle is well adapted to fluctuating discharges as a function of natural seasonality (Doadrio and Aldeguer, 2007). Moreover, the stream flow is one of the main drivers of the fish population dynamics, as it is demonstrated in Mediterranean rivers and elsewhere (Alonso-González *et al.*, 2004; Lobón-Cerviá and Mortensen, 2005). As explained above, the exotic species are a strong hazard for native freshwater fish in Mediterranean areas (Doadrio, 2001b; Hermoso and Clavero, 2011; Smith and Darwall, 2006). The minimum monthly flow from April to June (MinMF) is important to *L. guiraonis* because this species migrates in these months for spawning (Kottelat and Freyhof, 2007). IBMWP, as a biological indicator of the water quality, was found as a fundamental variable to predict fish distribution in Mediterranean rivers against other water quality indices (Carballo *et al.*, 2009; Olaya-Marín *et al.*, 2012).

The presence/absence model provides important information about the relation of *L. guiraonis* with biotic and habitat variables. This knowledge complements other models performed at the fish community level, and could be used to support future studies and practical decisions about the management and conservation of this species in the Júcar River Basin District. The density model did not permit to establish a clear relationship between the predictive variables and density of *L. guiraonis*, which should be considered in future studies to advance in the understanding of ecological interactions in Mediterranean ecosystems, and the critical factors for the species population enhancement.

The application of multilayer perceptron artificial neural networks had shown in several studies their capability to model complex ecological patterns and processes, with higher performances than traditional statistical approaches (Franklin, 2010; Olden *et al.,* 2008). Moreover, with other techniques is difficult to represent dataset patterns and trends. Given the nonlinearities in ecological processes and patterns, linear modelling is not a promising field to develop predictive models. Despite the advantages of ANN, they have been categorised as *black-box* models due to the little information given by the network about the relationship of each input variable and the dependent variables. this is explained by the fact that these relationships are implicit in the architecture of the MLP model. The black-box condition is the main disadvantage of MLP in contrast to traditional statistical approaches, in which, we could quantify the influence of each independent variable in the modelling process and the level of confidence in the prediction.

Nevertheless, several methods have been developed to overcome this issue of MLP; one of them is the partial derivative method (PaD), which is used to assess the contribution of each input variable in the prediction. PaD have been considered the most useful method to identify the degree of contribution of input variables in ANN models (Park and Chon, 2007), but PaD is calculated in relation to one independent variable at a time. Thus, when a predictive variable interacts with other one is difficult to explicitly represent the relationship, as it was observed in PaD analysis of the density model in this work. The improvement of the future models with the implementation of new techniques like PaD2, is clearly a line of research with promising results in the field of the data-driven modelling approaches.

Acknowledgements

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the projects SCARCE (Consolider-Ingenio 2010 CSD2009-00065) and POTECOL “Evaluación del Potencial Ecológico de Ríos Regulados por Embalses y Desarrollo de Criterios para su mejora según la Directiva Marco del Agua" (CGL2007-66412). In addition, the present research was developed in the frame of the EU-funded HolRiverMed project (Marie Curie Actions and in particular the Intra-European Fellowships for the mobility of European researchers). We thank the Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment) for the data provided to develop this study and we also owe our gratitude to Juan Jiménez for the collaboration in building the first fish database for this research.

REFERENCES

Aguilella A, Riera J, Gómez-Serrano MA, Mayoral O, Moreyra E. Evaluación del estado ecológico de los ríos de la cuenca hidrográfica del Júcar mediante el uso del índice QBR. Valencia: Jardí Botànic, Universitat de València; 2005.

Alba-Tercedor A. Macroinvertebrados acuaticos y calidad de las aguas de los ríos. Almería: IV Simposio del Agua en Andalucía (SIAGA); 1996.

Alba-Tercedor J, Sánchez-Ortega A. Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de Hellawell (1978). Limnetica 1988; 4: 51-56.

Almeida D, Grossman GD. Utility of direct observational methods for assessing competitive interactions between non-native and native freshwater fishes. Fisheries Manag. Ecol. 2012; 19: 157-166.

Alpaydın E. Introduction to Machine Learning, 2nd ed. Massachusetts: MIT Press; 2010.

Aparicio E, Carmona-Catot G, Moyle PB, García-Berthou E. Development and evaluation of a fish-based index to assess biological integrity of Mediterranean streams. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2011; 21: 324-337.

Baran P, Lek S, Delacoste M, Belaud A. Stochastic models that predict trout population density or biomass on a mesohabitat scale. Hydrobiologia 1996; 337: 1-9.

Baron JS, Poff NL, Angermeier PL, Dahm CN, Gleick PH, Hairston NG, *et al.* Meeting ecological and societal needs for freshwater. Ecol. Appl. 2002; 12: 1247-1260.

Bernardo JM, Ilhéu M, Matono P, Costa AM. Interannual variation of fish assemblage structure in a Mediterranean river: implications of streamflow on the dominance of native or exotic species. River Res. Appl. 2003; 19: 521-532.

Bishop CM. Neural Networks for Pattern Recognition. Oxford: University Press; 1996.

Bonada N, DolÉDec S, Statzner B. Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: implications for future climatic scenarios. Global Change Biology 2007; 13: 1658-1671.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, *et al*. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob. Ecol. Biogeogr. 2012; 21: 481-497.

Brosse S, Arbuckle CJ, Townsend CR. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. Biodivers. Conserv. 2003; 12: 2057-2075.

Brosse S, Guegan J-F, Tourenq J-N, Lek S. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 1999; 120: 299-311.

Brosse S, Lek S. Modelling roach (Rutilus rutilus) microhabitat using linear and nonlinear techniques. Freshw. Biol. 2000; 44: 441-452.

Brown GW, Krygier JT. Effects of Clear-Cutting on Stream Temperature. Water Resour. Res. 1970; 6: 1133-1139.

Buckland ST, Elston DA. Empirical Models for the Spatial Distribution of Wildlife. Journal of Applied Ecology 1993; 30: 478-495.

Caissie D. River discharge and channel width relationships for New Brunswick rivers. Canadian Technical Report of Fisheries and Aquatic Sciences, 2637, 2006a, pp. 26.

Caissie D. The thermal regime of rivers: a review. Freshw. Biol. 2006b; 51: 1389-1406.

Caissie D, El-Jabi N. Comparison and regionalization of hydrologically based instream flow techniques in Atlantic Canada. Can. J. Civ. Eng. 1995; 22: 235-246.

Carballo R, Cancela J, Iglesias G, Marín A, Neira X, Cuesta T. WFD indicators and definition of the ecological status of rivers. Water Resour. Manag. 2009; 23: 2231-2247.

Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, Foley JA, *et al.* Ecological Forecasts: An Emerging Imperative. Science 2001; 293: 657-660.

Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960; 20: 37-46.

Corbacho C, Sánchez JM. Patterns of species richness and introduced species in native freshwater fish faunas of a Mediterranean-type basin: the Guadiana River (southwest Iberian Peninsula). Regul. River. 2001; 17: 699-707.

Costa RMS, Martínez-Capel F, Muñoz-Mas R, Alcaraz-Hernández JD, Garófano-Gómez V. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, Parachondrostoma arrigonis (river Cabriel, Spain). River Res. Appl. 2012; 28: 740-752.

CHJ. Explotación de la red de vigilancia de la calidad de las aguas, mediante índices bióticos, en el ámbito de la Confederación Hidrográfica del Júcar Valencia: Confederación Hidrográfica del Júcar; 2008.

D'Heygere T, Goethals PLM, De Pauw N. Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. Ecol. Model. 2006; 195: 20-29.

Davis A, Jenkinson L, Lawton J, Shorrocks B, Wood S. Making mistakes when predicting shifts in species range in response to global warming. Nature 1998; 391: 783-786.

De Sostoa A. Las comunidades de peces en las cuencas mediterráneas: Caracterización y Problemática In: Doadrio I, editor. Atlas y Libro Rojo de los Peces Continentales de España. CSIC y Ministerio del Medio Ambiente, Madrid, 2002, pp. 51-56.

Demuth H, Beale M, Hagan M. Neural Networks toolbox 6. Users Guide. Matlab. Natick, Massachusetts The MathWorks, Inc.; 2010.

Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 1995; 2: 1-4.

Doadrio I. Atlas y Libro Rojo de los Peces Continentales de España. Madrid: Museo Nacional de Ciencias Naturales; 2001.

Doadrio I, Aldeguer M. La invasión de especies exóticas en los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Drew CA, Wiersma Y, Huettmann F. Predictive species and habitat modeling in landscape ecology: concepts and applications. New York: Springer; 2011.

Estrela T, Fidalgo A, Fullana J, Maestu J, Pérez MA, Pujante AM. Júcar Pilot River Basin, provisional article 5 report Pursuant to the Water Framework Directive. Valencia: Confederación Hidrográfica del Júcar; 2004.

Evans JS, Cushman S. Gradient modeling of conifer species using random forests. Landsc. Ecol. 2009; 24: 673-683.

Feio MJ, Poquet JM. Predictive Models for Freshwater Biological Assessment: Statistical Approaches, Biological Elements and the Iberian Peninsula Experience: A Review. International Review of Hydrobiology 2011; 96: 321-346.

Ferreira T, Oliveira J, Caiola N, De Sostoa A, Casals F, Cortes R, *et al.* Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Manag. Ecol. 2007; 14: 473-481.

Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 1997; 24: 38-49.

Filipe AF, Filomena Magalhães M, Collares-Pereira MJ. Native and introduced fish species richness in Mediterranean streams: the role of multiple landscape influences. Divers. Distrib. 2010; 16: 773-785.

Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Glob. Ecol. Biogeogr. 2007; 16: 24-33.

Franklin J. Mapping species distributions: spatial inference and prediction. New York: Cambridge University Press; 2010.

Freyhof J, Brooks E. European Red List of Freshwater Fishes. Luxembourg: Publications Office of the European Union; 2011.

Gabriels W, Goethals P, Dedecker A, Lek S, De Pauw N. Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. Aquat. Ecol. 2007; 41: 427-441.

García de Jalón D, González del Tánago M. Obras hidráulicas y ecosistemas fluviales. Madrid: EOI; 2007.

García de Jalón D, Sánchez Navarro R, Serrano J. Alteraciones de los regímenes de caudales de los ríos. Madrid: Ministerio de Medio Ambiente; 2007.

Garófano-Gómez V, Martínez-Capel F, Peredo-Parada M, Olaya-Marín EJ, Muñoz-Mas R, Costa R, *et al.* Assessing hydromorphological and floristic patterns along a regulated Mediterranean river: The Serpis River (Spain). Limnetica 2011; 30: 307-238.

Gasith A, Resh VH. Streams in mediterranean climate regions: Abiotic Influences and Biotic Responses to Predictable Seasonal Events. Annu. Rev. Ecol. Syst. 1999; 30: 51-81.

Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 2003; 160: 249-264.

Gevrey M, Dimopoulos I, Sovan L. Two-way interaction of input variables in the sensitivity analysis of neural network models. Ecol. Model. 2006; 195: 43-50.

Goethals P, Dedecker A, Gabriels W, Lek S, De Pauw N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 2007; 41: 491-508.

Gortázar J, Alonso González C, Iturriaga C, Hernández D, Baeza Sanz D, García de Jalón D. Estudio hidrobiológico de la cuenca del río Cabriel en las provincias de Albacete y Cuenca. Madrid: Castilla la Mancha (Ecohidraúlica SL-FUCOVASA) 2007.

Granado-Lorencio C. Ecología de peces. Sevilla: Universidad de Sevilla; 1996.

Granado-Lorencio C. Ecología de comunidades: el paradigma de los peces de agua dulce. Sevilla: Universidad de Sevilla; 2000.

Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecology Letters 2005; 8: 993-1009.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristics curve. Radiology 1982; 143: 29-36.

Hauser-Davis RA, Oliveira TF, Silveira AM, Silva TB, Ziolli RL. Case study: Comparing the use of nonlinear discriminating analysis and Artificial Neural Networks in the classification of three fish species: acaras (Geophagus brasiliensis), tilapias (Tilapia rendalli) and mullets (Mugil liza). Ecol. Inform. 2010; 5: 474-478.

Hellawell JM. Biological surveillance of rivers. Stevenage: Water Research Center; 1978.

Hermoso V, Clavero M. Threatening processes and conservation management of endemic freshwater fish in the Mediterranean basin: a review. Mar. Freshwater Res. 2011; 62: 244-254.

Hrachowitz M, Soulsby C, Imholt C, Malcolm IA, Tetzlaff D. Thermal regimes in a large upland salmon river: a simple model to identify the influence of landscape controls and climate change on maximum temperatures. Hydrol. Process. 2010; 24: 3374-3391.

Hsieh WW. Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels. Vancouver: Cambridge University Press; 2009.

Ibarra AA, Gevrey M, Park Y-S, Lim P, Lek S. Modelling the factors that influence fish guilds composition using a back-propagation network: assessment of metrics for indices of biotic integrity. Ecol. Model. 2003; 160: 281-290.

Isa IS, Omar S, Saad Z, Osman MK. Performance comparison of different multilayer perceptron network activation functions in automated weather classification. Proceedings of the 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation. Kota Kinabalu, Malaysia; 2010. p. 71-75.

Isaak DJ, Wollrab S, Horan D, Chandler G. Climate change effects on stream and river temperatures across the northwest U.S. from 1980–2009 and implications for salmonid fishes. Climatic Change 2012; 113: 499-524.

IUCN. IUCN Red List of Threatened Species. Version 2012.1;www.iucnredlist.org (downloaded on 14 September 2012), 2012.

Jackson DA, Peres-Neto PR, Olden JD. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 2001; 58: 157–170.

Jiménez J, Lacomba I, Sancho V, Risueño P. Peces continentales, anfibios y reptiles de la Comunidad Valenciana. Valencia: Generalitat Valenciana; 2002.

Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, Teachers and Scientists. Berlin: Springer-Verlag; 2011.

Jorgensen SE, Fath BD. Fundamentals of ecological modelling: applications in environmental management and research. 4th ed. Amsterdam: Elsevier; 2011.

Joy MK, Death RG. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. Freshw. Biol. 2004; 49: 1036-1052.

Koch GG, Landis JR, Freeman JL, Freeman DH, Jr., Lehnen RG. A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data. Biometrics 1977; 33: 133-158.

Kottelat M, Freyhof J. Handbook of European freshwater fishes. Berlin, Germany: Kottelat, Cornol, Switzerland and Freyhof; 2007.

Kurková V. Kolmogorov's theorem and multilayer neural networks. Neural Netw. 1992; 5: 501-506.

Lake PS, Bond N, Reich P. Linking ecological theory with stream restoration. Freshw. Biol. 2007; 52: 597-615.

Leclere J, Oberdorff T, Belliard J, Leprieur F. A comparison of modeling techniques to predict juvenile 0+ fish species occurrences in a large river system. Ecol. Inform. 2011; 6: 276-285.

Lek S, Scardi M, Verdonschot P, Descy JP, Park YS, editors. Modelling community structure in freshwater ecosystems. Berlin: Springer-Verlag; 2005.

Leopold LB, Maddock T. The hydraulic geometry of stream channels and some physiographic implications. Washington: U.S. Govt. Print. Off.; 1953.

Leopold LB, Wolman MG, Miller JP. Fluvial processes in geomorphology. San Francisco: W.H. Freeman; 1964.

Maceda-Veiga A, De Sostoa A. Observational evidence of the sensitivity of some fish species to environmental stressors in Mediterranean rivers. Ecol. Indic. 2011; 11: 311-317.

Magnuson JJ, Crowder LB, Medvick PA. Temperature as an Ecological Resource. American Zoologist 1979; 19: 331-343.

Manel S, Williams HC, Ormerod SJ. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 2001; 38: 921-931.

Martínez-Capel F, Costa R, Muñoz-Mas R. Evaluación de las poblaciones de peces en el río Júcar bajo el embalse de Alarcón, en las comarcas de la Manchuela Conquense (Cuenca) y la Mancha Júcar-centro (Albacete). Valencia: Universitat Politecnica de Valencia; 2008.

Martínez-Capel F, GarcÍa de Jalón D, Werenitzky D, Baeza D, Rodilla-AlamÁ M. Microhabitat use by three endemic Iberian cyprinids in Mediterranean rivers (Tagus River Basin, Spain). Fisheries Manag. Ecol. 2009; 16: 52-60.

Martínez-Muro. Control de calidad de aguas fluviales: diseño y puesta a punto de indicadores físico-químicos y biológicos. València: Universitat de València, Facultat de Ciències Químiques; 2003.

Mas-Martí E, García-Berthou E, Sabater S, Tomanova S, Muñoz I. Comparing fish assemblages and trophic ecology of permanent and intermittent reaches in a Mediterranean stream. Hydrobiologia 2010; 657: 167-180.

Mastrorillo S, Lek S, Dauba F, Belaud A. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshw. Biol. 1997; 38: 237-246.

Munné A, Prat N, Solà C, Bonada N, Rieradevall M. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2003; 13: 147-163.

Muñoz-Mas R, Martínez-Capel F, Schneider M, Mouton AM. 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. 2012; 440: 123.

Oberdorff T, Guégan J-F, Hugueny B. Global scale patterns of fish species richness in rivers. Ecography 1995; 18: 345-352.

Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernández JD. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). Sci. Total Environ. 2012; 440: 95-105.

Olden JD, Jackson DA. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 2002; 154: 135-150.

Olden JD, Lawler JJ, Poff NL. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Özesmi SL, Tan CO, Özesmi U. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 2006; 195: 83-93.

Park Y-S, Lek S, Chon T-S, Verdonschot FM. Evaluation of Environmental Factors to Determine the Distribution of Functional Feeding Groups of Benthic Macroinvertebrates Using an Artificial Neural Network. Journal of Ecology and Field Biology 2008; 31: 233-241.

Park YS, Chon TS. Biologically-inspired machine learning implemented to ecological informatics. Ecol. Model. 2007; 203: 1-7.

Pearce J, Ferrier S. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 2000; 133: 225-245.

Piekniewski F, Rybicki L. Visual comparison of performance for different activation functions in MLP networks. Proceedings of International Joint Conference on Neural Networks: IJCNN '04; 2004. p. 2947-2952.

Prchalová M, Slavík O, Bartoš L. Patterns of cyprinid migration through a fishway in relation to light, water temperature and fish circling behaviour. International Journal of River Basin Management 2006; 4: 213-218.

Provencher M, Lamontagne MP. Méthode de la détermination d'un indice d'appréciation de la qualité des aux selon différentes utilisations. Québec: Service de la qualité des eaux. Ministère des Richesses naturelles; 1977.

Qin Z, Su G-l, Zhang J-e, Ouyang Y, Yu Q, Li J. Identification of important factors for water vapor flux and CO_2 exchange in a cropland. Ecol. Model. 2010; 221: 575-581.

Ryan M, Müller C, Di HJ, Cameron KC. The use of artificial neural networks (ANNs) to simulate N_2O emissions from a temperate grassland ecosystem. Ecol. Model. 2004; 175: 189-194.

Sabater S, Tockner K. Effects of Hydrologic Alterations on the Ecological Quality of River Ecosystems Water Scarcity in the Mediterranean. In: Sabater S, Barceló D, editors. 8. Springer Berlin / Heidelberg, 2010, pp. 15-39.

Schlosser IJ. Stream fish ecology: a landscape perspective. Bioscience 1991; 41: 704-712.

Smith KG, Darwall WRT, editors. The status and distribution of freshwater fish endemic to the mediterranean basin. Gland, Switzerland/Cambridge, UK.: IUCN -The World Conservation Union; 2006.

Swets J. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285-1293.

Tirelli T, Pessani D. Use of decision tree and artificial neural network approaches to model presence/absence of *Telestes muticellus* in piedmont (North-Western Italy). River Res. Appl. 2009; 25: 1001-1012.

Tirelli T, Pessani D. Importance of feature selection in decision-tree and artificial-neural-network ecological applications. Alburnus alburnus alborella: A practical example. Ecol. Inform. 2011; 6: 309-315.

Tirelli T, Pozzi L, Pessani D. Use of different approaches to model presence/absence of *Salmo marmoratus* in Piedmont (Northwestern Italy). Ecol. Inform. 2009; 4: 234-242.

Vezza P, Parasiewicz P, Rosso M, Comoglio C. Defining minimum environmental flows at regional scale: application of mesoscale habitat models and catchments classification. River Res. Appl. 2012; 28: 675-792.

Vila-Gispert A, Alcaraz C, García-Berthou E. Life-history traits of invasive fish in small Mediterranean streams. Biol. Invasions 2005; 7: 107-116-116.

Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F. Recent advances in stream and river temperature research. Hydrol. Process. 2008; 22: 902-918.

Zarkami R, Sadeghi R, Goethals P. Use of fish distribution modelling for river management. Ecol. Model. 2012; 230: 44-49.

Zhang G, Eddy Patuwo B, Y. Hu M. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 1998; 14: 35-62.

Chapter 6

Conclusion and future work

6.1 General conclusions

In the following paragraphs I have summarized the conclusions of the PhD thesis, according to the general and specific objectives mentioned in previous chapters.

The ANN model built to predict native fish richness combines variables describing physical habitat and water quality, and it has demonstrated potential to identify the primary drivers of fish species richness patterns in Mediterranean rivers. The most critical variables at the basin scale were the index of water quality based on invertebrates (IBMWP), the proportion of riffle habitat, and the mean annual flow. Simulating the effect of obsolete weirs removal in the Júcar River, the model indicated a significant rise of native fish richness in response to the increase in channel length without artificial barriers and the potential increase in riffle proportion.

Based on future expectations about mean annual flow reduction in the Júcar River Basin (related to climate change), a simulation of the potential effects indicated a decrease of fish species richness. Due to this potential degradation of the ecological status, and the expected reduction of discharges, it is very important to adapt the water management strategies to address the consequences of the global change. This ANN model at basin-scale means the first step for modelling fish communities in more complex simulations, at different spatial and time scales, in order to assess the effects of water scarcity and global change on the Mediterranean fish communities.

Based on the finding of this PhD thesis, the habitat alterations are the main hazardous factors for the conservation of the native fish species in the Júcar, Cabriel, and Turia rivers. This study has shown that habitat alterations (including reduction of connectivity) are more important than biotic interactions (e.g. related to invasive species) to predict native fish richness. The results suggest that a Mediterranean fluvial system subject to anthropic disturbances and regulations is more vulnerable to exotic species invasion than a natural system, being the habitat degradation the driver of the ecological decline. The analysis based on the ANN models provides ideas to improve the ecological status in the Mediterranean freshwater ecosystems, and also to prevent the loss of biodiversity and ecological integrity of the fluvial ecosystem. Examples of these measures are the removal of abandoned weirs, design and operation of optimal environmental flow regimes, water quality improvement, and enhancement of the natural riparian vegetation. However, these prioritized actions must be carefully interpreted in the context of each river basin, where a different ranking of the environmental controls is possible, and the hierarchy of scales in the habitat factors may produce different effects.

Artificial neural networks and Random Forests constitute valuable tools to predict fish richness; their comparison showed that the best predictive method cannot be chosen a priori. Looking at the results, we could state that the use of more than one ML technique on the same study area was helpful, not only to identify the best model, but also to interpret the goodness of the results. ANN and RF models found the proportion of hydro-morphological units (HMU) as important variables to predict fish richness; particularly, percentage of riffles (RIF) was selected in ANN, and percentage of runs (RUN) and rapids (RAP) in RF. Even if the two modelling approaches arrived to identify different HMU types as predictive variables, one can see that the spatial distribution and dynamics of mesohabitats are important to model native species richness. As operational procedure for future studies on fish species richness, the comparison of different ML methods may assist the critical analysis of the results; it is also recommended to repeat the analysis in other study sites within the Mediterranean regions.

The presence/absence model provides important information about the relation of *Luciobarbus guiraonis* with biotic and habitat variables; this new knowledge could be used to support future studies and practical decisions about the management and conservation of this species in the Júcar, Cabriel and Turia rivers. Density model indicates that PaD method does not permit to establish a clear relationship between each predictive variable and the species density. This could be due to an interaction between FCV with PEF or another variable. For this reason, it is recommended to explore in future analysis the application of PaD2 analysis to depict the contribution of all possible pair-wise combinations of input variables.

6.2 FUTURE RESEARCH

Based on the findings of this thesis, the following research lines are proposed:

- **Elucidating the patterns and drivers of freshwater fish invasions:** Are invasive species a symptom of habitat modification and degradation, or are they drivers of diversity loss? Is removing invasive species crucial to prevent the extinction of native fish species, or is it largely a waste of time and resources? There are few studies concerning the relationship of invasive species and ecosystem change (Gurevitch and Padilla, 2004; Hermoso *et al.,* 2011; Spieles, 2010). Although the rapid growth of the invasions is associated with the decline of fish communities, it is not sufficient to state a causal relationship (i.e., association does not imply causation) (Spieles, 2010). This question was analysed in deep in this thesis, but it still needs further research to build new theories supporting the decision-making in river restoration (Gurevitch and Padilla, 2004; Olden *et al.*, 2010; Spieles, 2010).

- **Understanding the interactive effects of multiple stressors in freshwater ecosystems.** There are some evidences indicating that the synergistic effects among stressors (e.g. habitat loss, species invasions, pollution effects on fish, overharvesting and climate change) may accelerate the extinction process of freshwater fishes (Leprieur *et al.,* 2009; Olden *et al.,* 2010). 'Synergistic' means the simultaneous effect of separate processes with a superior total effect than the sum of individual effects (Leprieur *et al.,* 2009). Field observations and their analysis have shown that the above impacts can individually increase the risk of fish extinction in freshwater ecosystems (Dudgeon *et al.,* 2006). Our current understanding of the interactive effects is undeveloped (Olden *et al.,* 2010), it is of a great importance because an on-going extinction crisis is affecting Europe's freshwater fishes, and ambitious conservation actions (including the adequate protection and management of key freshwater habitats) are urgently needed (Freyhof and Brooks, 2011). An important field of research will be the incorporation of multiple drivers of environmental changes to analyse their synergistic effects on freshwater ecological processes (Olden *et al.,* 2010). Sensitivity of fish density as a function of habitat degradation is unclear; therefore, it will be interesting to analyse the synergistic effect of environmental change on other characteristics of the fish community like percentage of individuals with anomalies, age structure of native fish populations, abundance of native fishes, loss of native species, and alien fish pressure. These variables were identified by Aparicio *et al.* (2011) as key factors to define stream integrity.

- **Assessment of predictive uncertainty and study of the sources of uncertainty in aquatic ecosystem modelling.** Machine learning techniques have proved to be a useful approach to advance our understanding of ecological phenomena (Drew *et al.,* 2011; Olden *et al.,* 2008). However results will inevitably contain some degree of uncertainty (Jopp *et al.,* 2011; Peters *et al.,* 2009) and rarely has been taken into account when the robustness of ecological models was evaluated (Bruce G, 2012; Peters *et al.*, 2009).

REFERENCES

Aparicio E, Carmona-Catot G, Moyle PB, García-Berthou E. Development and evaluation of a fish-based index to assess biological integrity of Mediterranean streams. Aquat. Conserv.: Mar. Freshwat. Ecosyst. 2011; 21: 324-337.

Bruce G M. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecol. Model. 2012; 230: 50-62.

Drew CA, Wiersma Y, Huettmann F. Predictive species and habitat modeling in landscape ecology: concepts and applications. New York: Springer; 2011.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, *et al*. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 2006; 81: 163-182.

Freyhof J, Brooks E. European Red List of Freshwater Fishes. Luxembourg: Publications Office of the European Union; 2011.

Gurevitch J, Padilla DK. Are invasive species a major cause of extinctions? Trends Ecol. Evol. 2004; 19: 470-474.

Hermoso V, Clavero M, Blanco-Garrido F, Prenda J. Invasive species and habitat degradation in Iberian streams: an analysis of their role in freshwater fish diversity loss. Ecol. Appl. 2011; 21: 175-188.

Jopp F, Reuter H, Breckling B, editors. Modelling complex ecological dynamics: an Introduction into ecological modelling for students, Teachers and Scientists. Berlin: Springer-Verlag; 2011.

Leprieur F, Brosse S, García-Berthou E, Oberdorff T, Olden JD, Townsend CR. Scientific uncertainty and the assessment of risks posed by non-native freshwater fishes. Fish. Fish. 2009; 10: 88-97.

Olden JD, Kennard MJ, Leprieur F, Tedesco PA, Winemiller KO, García-Berthou E. Conservation biogeography of freshwater fishes: recent progress and future challenges. Divers. Distrib. 2010; 16: 496-513.

Olden JD, Lawler JJ, Poff NL. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 2008; 83: 171-193.

Peters J, Verhoest NEC, Samson R, Van Meirvenne M, Cockx L, De Baets B. Uncertainty propagation in vegetation distribution models based on ensemble classifiers. Ecol. Model. 2009; 220: 791-804.

Spieles DJ. Protected Land: Disturbance, Stress, and American Ecosystem Management Springer 2010.