HABITAT SUITABILITY MODELING WITH RANDOM FOREST AS A TOOL FOR FISH CONSERVATION IN MEDITERRANEAN RIVERS

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

FRANCISCO MARTINEZ-CAPEL
Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC)
Universitat Politècnica de València

RAFAEL MUÑOZ-MAS
Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC)
Universitat Politècnica de València

JUAN DIEGO ALCARAZ-HERNANDEZ
Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC)
Universitat Politècnica de València

CLAUDIO COMOGLIO
Department of Land, Environment and Geo-Engineering (DITAG)
Politecnico di Torino
24, C.so Duca degli Abruzzi 10129 Torino, Italy

Prediction of fish presence is needed in many branches of regulated river management, including the definition of environmental flows and habitat restoration measures for wildlife conservation. Research on river ecology has indicated that fish species distribution can be related to habitat attributes, and models with high predictive performance can be obtained by combining biotic and abiotic habitat descriptors. In four selected reference sites of the Cabriel River (province of Cuenca, Spain), the presence of Eastern Iberian barbel (*Luciobarbus guiraonis*) was related to environmental variables linked to different mesohabitat characteristics. By means of Random Forest (RF), the data collected in the field were used to predict fish presence for two key bioperiods: migration and spawning (April-June) and rearing and growth (July-September). The aims of this study are (i) to select the most important habitat attributes for the fish presence (ii) to evaluate how biotic interactions among fish species affect habitat use and (iii) to examine the feasibility of using RF in building habitat suitability models for fish. Random Forest provided an indicator of variables' importance and the most parsimonious model was selected to define the lowest number of variables to be surveyed for future model applications, e.g. habitat restoration measures and prediction of areas with high habitat suitability which should be conserved. The preliminary results of this research were discussed, as well as possible future developments, showing potentials and limitations of Random Forest in building habitat models for fish.

1 INTRODUCTION

In the context of the European Habitats (92/43/EEC) and Water Framework (2000/60/EEC) Directives, endemic and threatened fish species should be the targets of biodiversity safeguard and wildlife conservation actions (Hayer *et al.* [1]). Habitat suitability models have therefore a number of important applications for the conservation and management of target fish species (Mouton *et al.* [2]), including environmental flows assessment (Vezza *et al.* [3]) and habitat restoration measures (Costa *et al.* [4]). In particular, habitat models can be used (i) to predict species occurrence on the basis of habitat variables, (ii) to improve the understanding of species-habitat relationships and (iii) to quantify habitat requirements (see, Ahmadi-Nedushan *et al.* [5]).

When studying fish distribution, researchers make the assumption that the associations of fish species and habitat characteristics arise from either abiotic (physical and chemical habitat attributes) or biotic factors (e.g., biological interactions) or some combination of the two (Guisan *et al.* [6]). However, very few studies on habitat models explicitly include biotic factors for describing interactions among species (see for details, Elith *et al.* [7]). In freshwater ecology, and in particular for fish distribution analysis, meso-scale resolution (e.g., Vezza *et al.* [3]) can be used to capture the confounded effect of biotic and abiotic environmental variables, focusing on the ways in which mobile animals interact with the spatial arrangement of habitat characteristics (Addicott *et al.* [8]). Hydromorphological units – HMUs (or mesohabitats) has been increasingly used to describe and evaluate instream habitat structures (Parasiewicz [9], Gosselin *et al.* [10]) and the relevance of the mesoscale for fish studies was emphasized by Fausch *et al.* [11] because habitat features, such as cover sources or obstacles to fish movements, were best observed at this scale. Moreover, Mouton *et al.* [12] highlighted that future research in fish habitat modeling should take into account biotic interactions among species, which may play an important role in the habitat suitability assessment.

Most studies on fish–habitat relationships have focused on Salmonids because of their economic importance and ubiquity (Gosselin *et al.* [10]). This study is focused on the Eastern Iberian barbel (*Luciobarbus guiraonis*), a vulnerable fish species (Baillie *et al.* [13]) typical of the Mediterranean rivers of the Valencia region (i.e. between Mijares and Vinalopo, inclusive, Crivelli [14]). The fish population is declining due to the presence of water abstractions and habitat modification, which favour the alien species (Kottelat *et al.* [15], Doadrio [16]). Few studies have focused on the general ecology of the Eastern Iberian barbel (Crivelli [14]) and no habitat models are currently available in literature.

Recently, several studies (Cutler *et al.* [17], Kampichler *et al.* [18], Siroky [19]) have shown that, compared to other methodologies, RF models (Breiman [20]), a machine learning technique based on an automatic combination of decision trees, often reach top predictive performances in building predictive habitat models for species distribution. To develop a reliable and ecologically relevant species distribution model, in this research we used RF to predict the habitat suitability at meso-scale, based on combinations of physical and biological habitat descriptors. Two key bioperiods were considered: migration and spawning (April-June) and rearing and growth (July-September). The aims of the study are: (i) to select the most important habitat attributes for barbel presence (ii) to evaluate how biotic interactions among fish species affect habitat use and (iii) to examine the feasibility of using RF in building habitat suitability models for fish.

2 METHODS

2.1 Study area

This study was conducted in four selected sites of the Cabriel River (province of Cuenca, Spain), due to their reference habitat conditions (no or little human impact, *sensu* Vezza et al. [3]), natural flow regime and presence of *Luciobarbus guiraonis* (see Figure 1). The Cabriel River (220 km long and drainage area of 4750 km^2) is part of the Júcar River Basin, which is characterized by a typical Mediterranean climate (i.e. low flows and high evapotranspiration in summer and high flows in spring and autumn).

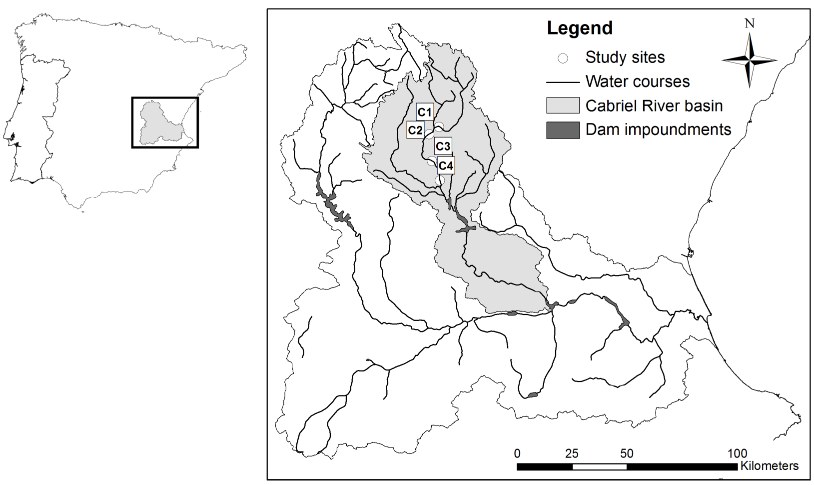

Figure 1. Location of the four study sites in the river Cabriel (Júcar River Basin - Eastern Spain). Main watercourses and dam impoundments are also reported in the map.

The mean elevation is equal to 1016 m a.s.l. (elevation ranges from 490 to 1790 m a.s.l.) and the mean annual precipitation is 500 mm. To describe reference habitat characteristics, the four study sites (named C1, C2, C3 and C4 in downstream order, see Figure 1) were located in the upper part of the Cabriel catchment, upstream of the large Contreras Dam (Costa *et al.* [4]). In this part of the catchment, the land cover (from the Corine Land Cover classification; Bossard *et al.* [21]) is mainly represented by forested areas (86%) and crops (12%). At C4, the median flow (Q50) is 2.74 m^3s^{-1}, high flow (Q5) is 15.83 m^3s^{-1} and low flow (Q95) is 0.94 m^3s^{-1}. Note that each site differs in terms of morphological characteristics (channel size, mean gradient, dominant substrates and cover sources), watershed area and flow duration curves (because of the presence of tributaries located between the sites). Thus, the four selected sites can be used to represent the different habitat features available for barbel in the Cabriel River.

2.2 Habitat description and fish data

The habitat survey described the changes in the HMUs characteristics with different flow rates. Habitat and fish population assessments were conducted within each site during 2006, 2007 and 2008, collecting data from a total amount of 240 mesohabitats. The survey design ensured equal sampling effort in the four selected reaches of the river and each study site was 1 ± 0.1 km long (usually longer to include complete HMUs, Costa *et al.* [4]). Most of the sites included five types of HMUs: pool, glide, rapid, riffle and run (Dolloff *et al.* [22]) . The surveys captured the instream habitat variability across the sites and over a range of discharges (between medium, Q40-Q50, and low flows, Q80-Q98). All the HMUs were identified and described following the procedure reported in Costa *et al.* [4]. In particular, for each HMU the following habitat attributes were collected (see Table 1): length, mean width, mean and maximum water depth, HMU gradient, types of substrate and cover sources.

Table 1. Code, description, unit and range of the habitat descriptors (i.e. biotic and abiotic parameters) included in the RF model.

Variable code	Description	Unit	Range
Wid	Mean channel width	m	2.7 - 20.0
Dmed	Mean water depth	m	0.29 - 3.52
Dmax	Maximum water depth	m	0.34 - 4.10
Vmed	Mean flow velocity	m/s	0.04 - 1.05
Grad	HMU Gradient	%	0.0 - 9.3
RK	Bedrock	%	0 - 100
CS	Coarse substrate (boulders and cobbles)	%	0 - 100
FS	Fine substrate (gravel and sand)	%	0 - 100
SC	Silt, clay and sludge	%	0 - 60
SI	Substrate Index	-	1 - 7
Veg	Submerged vegetation	%	0 - 90
Sh	Canopy shading	%	0 - 100
UB	Undercut banks	%	0 - 100
WD	Woody debris	no(0)/yes(1)	0 - 1
B	Boulders	no(0)/yes(1)	0 - 1
Wid150	Mean width of the 150 m stretch upstream the HMU	m	2.9 - 14.5
Dmed150	Mean depth of the 150 m stretch upstream the HMU	m	0.29 - 3.52
Dmax150	Maximum depth of the 150 m stretch upstream the HMU	m	0.34 - 4.10
Vmed150	Mean velocity of the 150 m stretch upstream the HMU	m/s	0.04 - 0.89
Grad150	Mean gradient of the 150 m stretch upstream the HMU	%	0.0 - 4.0
ACAC	Abundance of Southern Iberian chub (*Squalius pyrenaicus*)	no(0)/pres.(1)/ab.(2)	0 - 2
AJUC	Abundance of Jucar nase (*Parachondrostomas arrigonis*)	no(0)/pres.(1)/ab.(2)	0 - 2
ATAG	Abundance of Iberian straight-mouth nase (*Pseudochondrostoma polylepis*)	no(0)/pres.(1)/ab.(2)	0 - 2
AGOB	Abundance of Pyrenean gudgeon (*Gobio lozanoi*)	no(0)/pres.(1)/ab.(2)	0 - 2
ATRO	Abundance of brown trout (*Salmo trutta fario*)	no(0)/pres.(1)/ab.(2)	0 - 2

The mean flow velocity of each mesohabitat was derived by dividing the mean HMU cross-section by the value of the measured discharge. In order to capture the influence of upstream conditions on barbel mesohabitat use, channel width, mean and maximum water depth, mean flow velocity and mean gradient were also calculated for each HMU using the 150 m upstream stretch characteristics (Britton *et al.* [23]).

During the surveys, the fish were counted in each HMU by snorkeling, to consider their presence/absence during its diurnal routine and, at the same time, to avoid any damage to the target vulnerable species (Baillie *et al.* [13]). In particular, two divers conducted the underwater counts in three independent passes throughout each habitat unit (see Costa *et al.* [4] for details). This technique was chosen due to its representativeness of fish population densities at meso-scale (Gosselin *et al.* [10]) and the authors consider it was the most appropriate methodology for this study due to the morphological characteristics of the river (i.e. clear water, presence of deep pools and high density of riparian vegetation). To investigate the biological interactions among species (e.g., competition in habitat use), three classes of fish abundance, i.e. no (0), present (1) and abundant (2), were also added as biological mesohabitat attributes for all the fish species found in the Cabriel River (see Table 1).

2.3 Data analysis

The associations of mesohabitat characteristics with *Luciobarbus guiraonis* distribution was explored using Random Forest (RF) models (Cutler *et al.* [17], Breiman [20]) in order to establish habitat suitability criteria. Random Forest, as implemented in R (R Development Core Team 2009; Liaw *et al.* [24]) is an ensemble learning technique based on a combination of a large set of decision trees (CART, Breiman *et al.* [25]). Each tree is trained by selecting a random bootstrap subset X_i (i = bootstrap iteration which ranges from 1 to t, maximum number of trees) of the original dataset X and a random set of predictive variables (Liaw *et al.* [24]). As the response variable is categorical (fish presence/absence), we confine our attention to classification RF models. The algorithm for growing a random forest of t classification trees performs as follows (for full details see Breiman [20]):

t bootstrap samples X_i (training dataset) are randomly drawn with replacement from the original dataset, each containing approximately two third of the elements of the original dataset X. The elements not included in the training dataset are referred to 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.

For each bootstrap sample X_i, an unpruned classification tree is grown. At each node m variables are randomly selected and the best split is chosen among them.

The trees are fully grown and each is used to predict the OOB observations. In particular, the majority vote is taken by aggregating the predictions of the t trees and generate new out-of-bag data. Note that, because the OOB observations are not used in the fitting of the RF trees, the out-of-bag estimates are essentially cross-validated accuracy estimates.

Global RF accuracies and error rates (i.e. the OOB error, E_{OOB}, and the within-class errors, $E_{Class(j)}$) are computed using the out-of-bag predictions.

The E_{OOB} is also used to choose an optimal value of t and m. Firstly, 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 and variable interaction, with a large set of independent variables, is defined as [2*(t for E_{OOB} stabilization) = 5000] (Evans *et al.* [26]). Secondly, the m parameter (number of variables permutated at each node) is defined as the square root of the total number of predictor variables, with a minimum of m = 2 (Breiman [20]).

To assess the importance of each predictor variable, the values of the variable are randomly permuted for the OOB observations, and then the modified out-of-bag data are passed down the trees to get new predictions. The difference between the misclassification rate for the modified versus the original out-of-bag data, divided by the standard error, is a measure of the importance of the variable. A higher variable importance indicates a larger contribution to the RF prediction accuracy.

To identify the most parsimonious model we applied the Model Improvement Ratio (MIR) technique (details in, Murphy *et al.* [27]). To analyze the seasonal response to habitat changes, two different seasonal habitat models were developed, referring to migration and spawning (April-June) and rearing and growth (July-September) bioperiods (Kottelat *et al.* [15], Doadrio [16]). Lastly, the partial dependence plots provided a way to visualize the marginal effect of the selected independent variables on the predicted probabilities of barbel presence (details in Cutler *et al.* [17]).

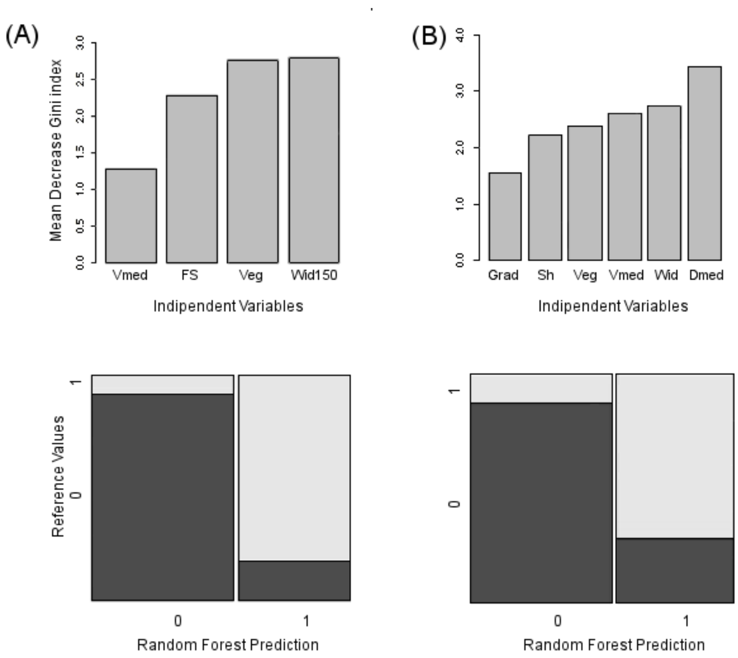

Figure 2. (A) Migration and spawning and (B) rearing and growth suitability models for Eastern Iberian barbel, built using only abiotic habitat descriptors. The most relevant variables and their relative importance by Mean Decrease Gini Index (Breiman [20]) are reported, along with the confusion matrixes of the selected RF models.

3 RESULTS

Figure 2 reports the migration and spawning and rearing and growth habitat models for Eastern Iberian barbel, built using only abiotic variables. According to the MIR technique the most parsimonious model was selected (see Murphy *et al.* [27] for details).

For barbel, shelters provided by submerged vegetation and flow velocity were selected as important habitat attributes for both analyzed bioperiods. During the migration and spawning, the model was also influenced by the fine substrate and the upstream conditions of channel width; during rearing and growth, mean water depth, channel width, canopy shading and HMU gradient were selected as relevant environmental variables.

Note that all predictive habitat models (see Table 2) were significant at $P<0.001$ and had high model-fit accuracy (ranging from 80 to 93%). Overall, kappa statistics are over 0.56 and models show high sensitivity/specificity values, indicating substantial predictions with low cross-classification error. In addition, the ROC area under curve (AUC) was over 0.80 in all cases and indicated good or excellent model performance. Although the models built using only physical habitat attributes performed well, one can observe how considering biotic interactions among species increases the global model performance (Table 2).

Table 2. RF models for Eastern Iberian barbel (for migration and spawning, M&S, and rearing and growth, R&G, bioperiods). Models were developed using only aboitic and both biotic and abiotic independent variables to investigate the influence of biotic interactions. The selected variables (in order of importance), model accuracy (%), sensitivity, specificity, Kappa (*k*), ROC area under curve (AUC) and significance (*P*) are reported for each model.

Model class	Bioperiod	Selected variables	Accuracy	Sensitivity	Specificity	*k*	AUC	*P*
Abiotic descriptors	M&S	Wid150, Veg, FS Vmed	90	0.89	0.88	0.74	0.93	<0.001
	R&G	Dmed, Wid, Vmed, Veg, Sh, Grad	80	0.78	0.82	0.56	0.80	<0.001
Biotic and abiotic descriptors	M&S	AGOB, FS, Veg, Wid150,	93	0.95	0.96	0.82	0.98	<0.001
	R&G	ACAC, AGOB, Dmed, Wid	85	0.86	0.84	0.66	0.88	<0.001

The abundance of Pyrenean gudgeon (*Gobio lozanoi*) seemed to negatively influence the barbel distribution, both in spring and summer seasons. Moreover, during Summer, also the abundance of Southern Iberian chub (*Squalius pyrenaicus*) demonstrates to influence the barbel presence in a negative sense, reflecting possible habitat competition among fish species.

To represent the marginal effect of a single variable included in the RF models, the partial dependence plots were used (Figure 3), showing the relationships between individual predictor variables and predicted probabilities of fish presence. For binary classification (i.e. presence/absence of fish), the y-axis on partial dependence plots is on the logit scale (see Cutler *et al.* [17]).

4 DISCUSSION

Elith *et al.* [7] outlined that more effective conservation of endangered species and aquatic biodiversity will require new approaches that recognize not only abiotic habitat parameters, but also the different biotic interactions among species. Many applications could benefit from these advances in modeling the ecological processes that shape species distributions (Guisan *et al.* [6]). According to Elith *et al.* [7], the aim of this paper is to gain further insight into habitat preferences of Easter Iberian barbel (*Luciobarbus guiraonis*) and its biotic interactions. By means of the RF technique we developed mesohabitat suitability criteria, using two reference bioperiods. The obtained models and the relevant variables can be important for prioritizing surveys and monitoring programmes (Poff *et al.* [28]), particularly for the definition of environmental flows and habitat restoration measures (e.g., Costa *et al.* [4]). In freshwater ecology, mesohabitats (or HMUs) can be considered the appropriate scale resolution to capture, from a fish species' viewpoint, the way in which mobile animals interact with the spatial arrangement of different habitat characteristics, also considering the seasonal habitat changes and migration behaviors (Fausch *et al.* [11]).

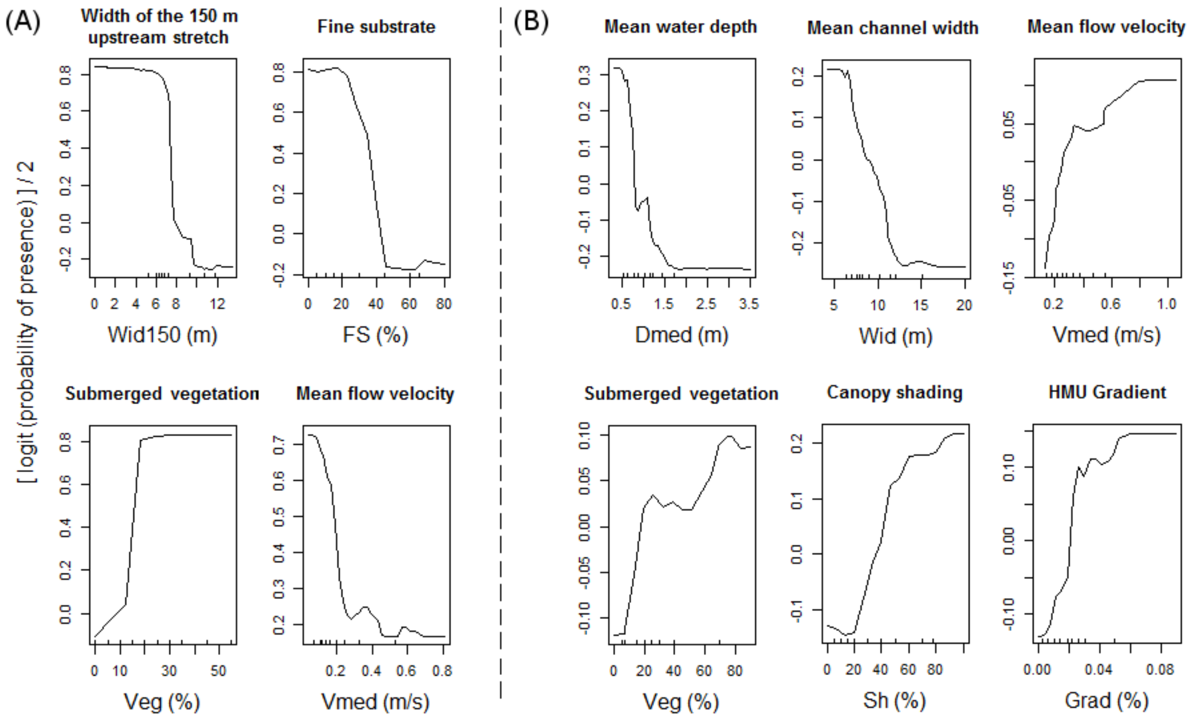

Figure 3. Partial dependence plots for (A) migration and spawning and (B) rearing and growth habitat suitability model for Eastern Iberian barbel. Partial plots represent the marginal effect of a single variable included in the RF model on the probability of fish presence, while averaging out the effect of all the other parameters.

In this research, the Eastern Iberian barbel presence was found to be related to various aspects of instream habitat, in relation to the two studied bioperiods (Table 2). The species occurrence during migration and spawning (M&G) was best predicted with variables describing the upstream channel size (Wid150), the amount of fine substrate (FS), shelters provided by submerged vegetation (Veg) and mean flow velocity (Vmed). Channel size, substrate and flow velocity are known to influence the occurrence of the species during upstream migrations (e.g., Kottelat *et al.* [15]), while the cover provided by submerged vegetation (also selected in the rearing and growth, R&G, model) can serve as resting and hiding area for the fish. In contrast, mean water depth, channel width and HMU gradient were chosen as the most important habitat attributes during R&G bioperiod. Moreover in the R&G model, the canopy shading was also selected as an important habitat characteristic, which is a finding consistent with the fish thermal requirements during summer (e.g., Vila-Gispert *et al.* [29]).

Partial dependence plot (Figure 3) showed a non-linear relationship between the logit of the probability of barbel presence and the selected predictor variables. It is interesting to note how, in the M&S model, the probability of presence drops rapidly with increasing flow velocity and then levels off; in contrast, for R&G bioperiod flow velocity has the opposite effect: the probability of presence increase and then levels off. These changings in habitat preferences are the likely reason for developing seasonal habitat models for fish and indicate the interesting potentials of the tree-based methods, and RF in particular, in this kind of analyses. Note that high velocity can be limiting during migration and low velocity during summer can be related to low discharge, little food availability and high water temperature.

As reported in Elvira [30] and Kottelat *et al.* [15] the population of Eastern Iberian barbel is declining due to habitat modification which favor alien species. In particular, the introduced species Pyrenean gudgeon (*Gobio lozanoi*) seems to compete in habitat selection with barbel, as gudgeon abundance has a negative influence on the probability of barbel presence (not showed). Moreover, during summer, also the abundance of Southern Iberian chub (*Squalius pyrenaicus*) demonstrates to influence the barbel presence, reflecting a possible increase in habitat competition during summer low flows.

The comparison between models developed using abiotic (only) and both biotic and abiotic parameters can be useful to evaluate if fish habitat selection is mainly driven (or not) by the instream physical characteristics (Mouton *et al.* [12]). In the case of the Cabriel River, the RF predictions (using only abiotic parameters) showed high accuracy (and moderate to high model specificity and sensitivity) and exhibited considerable promise in developing predictive models for fish conservation in Mediterranean rivers. However, modelling methods including also biotic interactions were considered more ecologically realistic to understand and describe the interplay of the different environmental variables for barbel distribution.

In several RF studies (Vincenzi *et al.* [31], He *et al.* [32]), the major interest is to identify the most important factors that affect the species distribution. In this research, the Model Improvement Ratio (MIR, Murphy *et al.* [27]) technique was applied in the RF variable selection, in order to optimize the parsimony of the model and identify the lowest number of variable to be surveyed for future model applications. Further research efforts (according to Olden *et al.* [33]) will be spent in comparing different statistical methods that best suit the characteristics of the data and best fit the proposed application, e.g. predicting potential sites for Eastern Iberian barbel habitat enhancement.

AKNOWLEDGMENT

This research was developed within the framework of the EU-funded HolRiverMed project (275577 - FP7-PEOPLE-2010-IEF, Marie Curie Actions).

REFERENCES

[1] Hayer CA, Wall SS, Berry Jr CR, "Evaluation of predicted fish distribution models for rare fish species in South Dakota", *North American Journal of Fisheries Management*, Vol. 28, No. 4, (2008), pp 1259-1269.

[2] 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", *Environmental Modelling & Software*, Vol. 26, No. 5, (2011), pp 615-622.

[3] Vezza P, Parasiewicz P, Rosso M, Comoglio C, "Defining minimum environmental flows at regional scale by using meso-scale habitat models and catchments classification", *River Research and Applications*, In press, (2011), pp

[4] 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 Research and Applications,* In press, (2011).

[5] Ahmadi-Nedushan B, St-Hilaire A, Bérubé M, Robichaud E, Thiémonge N, Bernard B, "A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment", *River Resesearch and Applications*, Vol. 22, (2006), pp 503-523.

[6] Guisan A, Thuiller W, "Predicting species distribution: offering more than simple habitat models", *Ecology Letters,* Vol. 8, No. 9, (2005), pp 993-1009.

[7] Elith J, Leathwick JR, "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time", *Annual Review of Ecology, Evolution, and Systematics*, Vol. 40, No. 1, (2009), pp 677-697.

[8] Addicott J, Aho J, Antolin M, Padilla D, Richardson J, Soluk D, "Ecological neighborhoods: scaling environmental patterns", *Oikos*, Vol. 49, (1987), pp 340-346.

[9] Parasiewicz P, "The MesoHABSIM model revisited", *River Research and Applications*, Vol. 23, No. 8, (2007), pp 893-903.

[10] Gosselin M, Petts G, Maddock I, "Mesohabitat use by bullhead (Cottus gobio)", *Hydrobiologia*, Vol. 652, (2010), pp 299-310.

[11] Fausch KD, Torgersen CE, Baxter CV, Li HW, "Landscapes to Riverscapes: Bridging the Gap between Research and Conservation of Stream Fishes", *BioScience*, Vol. 52, No. 6, (2002), pp 483-498.

[12] Mouton AM, Schneider M, Depestele J, Goethals PLM, De Pauw N, "Fish habitat modelling as a tool for river management", *Ecological Engineering*, Vol. 29, No. 3, (2007), pp 305-315.

[13] Baillie JEM, Hilton-Taylor C, Stuart SN, " *IUCN Red List of Threatened Species. A Global Species Assessment"*, IUCN, Gland, Switzerland; (2004).

[14] Crivelli AJ, "*The freshwater fish endemic to the Mediterranean region. An action plan for their conservation."*, Tour du Valat Publication, 171 p.; (1996).

[15] Kottelat M, Freyhof J, "*Handbook of European freshwater fishes"*, Publications Kottelat, Cornol, Switzerland. 646 p.; (2007).

[16] Doadrio I, "*Atlas y Libro Rojo de los Peces Continentales de Espana"*, 2001).

[17] Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ, "Random forests for classification in ecology", *Ecology*, Vol. 88, No. 11, (2007), pp 2783-2792.

[18] Kampichler C, Wieland R, Calmé S, Weissenberger H, Arriaga-Weiss S, "Classification in conservation biology: A comparison of five machine-learning methods", *Ecological Informatics*, Vol. 5, No. 6, (2010), pp 441-450.

[19] Siroky DS, "Navigating Random Forests and related advances in algorithmic modeling", *Statististical Surveys*, Vol. 3, No. (2009), pp 147-163.

[20] Breiman L, "Random Forest", *Machine Learning*, Vol. 45, (2001), pp 5-32.

[21] Bossard M, Feranec J, Otahel J, "*CORINE land cover technical guide – Addendum 2000"*, 2000).

[22] Dolloff CA, Hankin DG, G.H. R, "*Basinwide estimation of habitat and fish populations in streams"*, Gen. Tech. Rep. SE-GTR-83; (1993).

[23] Britton JR, Pegg J, "Ecology of European Barbel *Barbus barbus*: Implications for River, Fishery, and Conservation Management", *Reviews in Fisheries Science*, Vol. 19, No. 4, (2011), pp 321-330.

[24] Liaw A, Wiener M, "Classification and regression by Random Forest", *R News*, Vol. 2, (2002), pp 18-22.

[25] Breiman L, Friedman JH, Olshen R, Stone CJ, "*Classification and Regression Trees"*, 1984).

[26] Evans J, Cushman S, "Gradient modeling of conifer species using random forests", *Landscape Ecology*, Vol. 24, No. 5, (2009), pp 673-683.

[27] Murphy MA, Evans JS, Storfer A, "Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics", *Ecology*, Vol. 91, No. 1, (2010), pp 252-261.

[28] Poff NL, Zimmerman JKH, "Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows", *Freshwater Biology*, Vol. 55, (2010), pp 194-205.

[29] Vila-Gispert A, García-Berthou E, Moreno-Amich R, "Fish zonation in a Mediterranean stream: Effects of human disturbances", *Aquatic Sciences - Research Across Boundaries*, Vol. 64, No. 2, (2002), pp 163-170.

[30] Elvira B, "*Endangered freshwater fish of Spain."*, Basel: Birkhauser Verlag; (1996).

[31] Vincenzi S, Zucchetta M, Franzoi P, Pellizzato M, Pranovi F, De Leo GA, Torricelli P, "Application of a Random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy", *Ecological Modelling*, Vol. 222, No. 8, (2011), pp 1471-1478.

[32] He Y, Wang J, Lek-Ang S, Lek S, "Predicting assemblages and species richness of endemic fish in the upper Yangtze River", *Science of The Total Environment*, Vol. 408, No. 19, (2010), pp 4211-4220.

[33] Olden JD, Jackson DA, "A comparison of statistical approaches for modelling fish species distributions", *Freshwater Biology*, Vol. 47, No. 10, (2002), pp 1976-1995.