Aquat Sci
DOI 10.1007/s00027-013-0306-7

Aquatic Sciences

RESEARCH ARTICLE

Modelling habitat requirements of bullhead (*Cottus gobio*) in Alpine streams

Paolo Vezza · P. Parasiewicz · O. Calles · M. Spairani · C. Comoglio

Received: 5 February 2013 / Accepted: 27 August 2013

Abstract In the context of water resources planning and management, the prediction of fish distribution related to habitat characteristics is fundamental for the definition of environmental flows and habitat restoration measures. In particular, threatened and endemic fish species should be the targets of biodiversity safeguard and wildlife conservation actions. The recently developed meso-scale habitat model (MesoHABSIM) can provide solutions in this sense by using multivariate statistical techniques to predict fish species distribution and to define habitat suitability criteria. In this research, Random Forests (RF) and Logistic Regressions (LR) models were used to predict the distribution of bullhead (*Cottus gobio*) as a function of habitat conditions. In ten reference streams of the Alps (NW Italy), 95 mesohabitats were sampled for hydro-morphologic and biological parameters, and RF and LR were used to distinguish between absence/presence and presence/abundance of fish. The obtained models were compared on the basis of their performances (model accuracy, sensitivity, specificity, Cohen's kappa and area under ROC curve). Results indicate that RF outperformed LR, for both absence/presence (RF: 84 % accuracy, $k = 0.58$ and AUC $= 0.88$; LR: 78 % accuracy, $k = 0.54$ and AUC $=$ 0.85) and presence/abundance models (RF: 79 % accuracy, $k = 0.57$ and AUC $= 0.87$; LR: 69 % accuracy, $k = 0.43$ and AUC $= 0.81$). The most important variables, selected in each model, are discussed and compared to the available literature. Lastly, results from models' application in regulated sites are presented to show the possible use of RF in predicting habitat availability for fish in Alpine streams.

Keywords Mesohabitat · MesoHABSIM · Alpine streams · Cottus gobio · Habitat suitability

P. Vezza (✉)
Institut d'Investigació per a la Gestió Integrada de Zones Costaneres, Universitat Politècnica de València, C/Paranimf, 1, 46730 Grau de Gandia (Valencia), Spain
e-mail: paovez@upv.es

P. Vezza · O. Calles · C. Comoglio
Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Turin, Italy

P. Parasiewicz
Rushing Rivers Institute, Amherst, MA, USA

P. Parasiewicz
S. Sakowicz Inland Fisheries Institute, Zabieniec, Poland

O. Calles
Department of Biology, Karlstad University, Karlstad, Sweden

M. Spairani
FLUME s.r.l, Aosta, Italy

Introduction

To meet the aims set by the European Habitats (92/43/EEC) and Water Framework (2000/60/EEC) Directives, mitigation of anthropogenic hydro-morphological alterations affecting river ecosystems needs to be addressed. Therefore, endemic and threatened fish species should be the targets of habitat enhancement and river restoration actions (Hayer et al. 2008; Acreman and Ferguson 2010). Focusing on Italian rivers, water abstractions, habitat alterations and pollution are the three most important factors causing endangerment of aquatic biota (Crivelli 1996; Zerunian 2007; Regione Piemonte 2010) and creating the need for the conservation of fish populations and communities. In the last two decades, freshwater fish conservation has been increasingly considered; several works have

Published online: 12 September 2013

contributed to understand the effects of human activities on aquatic organisms (e.g., Gandolfi et al. 1991; Zerunian 2002b; Ciuffardi and Bassani 2005; Comoglio et al. 2007; Pini Prato et al. 2011) and have underlined that many taxa are running the risk of extinction (Zerunian 2002a).

Habitat models can play an important role in predicting spatial and temporal patterns of species distribution (Olden et al. 2002), having a number of important applications for the conservation of aquatic organisms (Ahmadi-Nedushan et al. 2006). In particular, statistical models allow users to predict species occurrence on the basis of habitat variables, provide useful insights and understanding of species-habitat relationships (Mouton et al. 2007), and quantify habitat requirements for environmental flow assessment (Vezza et al. 2012a).

Meso-scale resolution of freshwater fish habitat (i.e. mesohabitats) has been widely used as a template for examining the habitat use by aquatic fauna (Parasiewicz 2001, 2007; Borsányi et al. 2003; Gosselin et al. 2010, 2012). Hydromorphological units (HMUs) and mesohabitats (e.g., series of pools, riffles and rapids), which commonly correspond in size and location (Bain and Knight 1996; Parasiewicz 2007; Hauer et al. 2010), refer to stream units mentioned in the habitat classification of Frissell et al. (1986) and are related to the concept of the functional habitat; i.e., areas where animals can be observed for a significant portion of their diurnal routine (Kemp et al. 1999). They physically reflect the interplay between hydraulics and riverbed topography and can be inferred by visual observation of surface flow characteristics and verified by water depth, flow velocity and substrate types (Gosselin et al. 2010). Moreover, meso-scale approaches can include a large range of habitat variables in biological models, which can allow the understanding of fish behavior at larger spatial scales (Jewitt et al. 2001). Models predicting the mesohabitat suitability can therefore guide the design of habitat enhancement and river rehabilitation actions using fish requirements as ecological reference (Parasiewicz et al. 2012b). The mesohabitat simulation model (MesoHABSIM) can be used to define reference conditions for fish aimed at maintaining the natural ecological integrity of rivers and streams. Furthermore, the MesoHABSIM application has been recently extended to describe instream habitat in mountainous watercourses characterized by coarse substrate, high gradient and complex morphology (Vezza et al. 2012a, b).

Classification procedures are among the most widely used statistical methods to predict species distribution and habitat suitability (Mouton et al. 2010) and a number of computational statistical techniques are now available (Ahmadi-Nedushan et al. 2006). Among the different available approaches for building multivariate habitat suitability models, we focused our attention on Random Forests (RF, Breiman 2001; Cutler et al. 2007) and Logistic Regressions (LR, Parasiewicz 2007; Tirelli et al. 2009) to predict species distribution and to define fish mesohabitat requirements (e.g., Mouton et al. 2011). RF is a machine-learning technique, which combines many decision trees (or CART, Breiman et al. 1984) to produce prediction outcomes. It is suitable for both classification and regression (Breiman 2001) and competitive with or superior to most available methods in the literature (Cutler et al. 2007; Siroky 2009; Kampichler et al. 2010). On the other hand, LR is a type of multiple regression analysis used for predicting a binary variable based on several predictor variables, which can be both categorical and continuous. LR is already implemented in the MesoHABSIM methodology, in which the absence/presence or presence/abundance of fish are modeled using a logistic function (Parasiewicz 2007).

In this study, we used RF and LR to predict mesohabitat suitability for fish and evaluated their application in the framework of the MesoHABSIM simulation system. We focused our attention on bullhead (*Cottus gobio*), a small bottom-dwelling fish that is considered vulnerable in Italy (Zerunian 2007) and should be the target of future habitat enhancement measures in Alpine streams and rivers (Regione Piemonte 2007). Structured populations of bullhead are indeed rare and mainly located in streams with no or negligible degrees of impact (Regione Valle d'Aosta 2008; Regione Piemonte 2010). Several studies have focused on the general ecology of bullhead across Europe (Roussel and Bardonnet 1996; Gosselin et al. 2010). However, no habitat models are currently available to describe habitat requirements of this fish species in Alpine environments and there is a lack of detailed knowledge on Italian subpopulations, which are threatened by habitat alteration and water abstraction (Regione Valle d'Aosta 2008; Regione Piemonte 2010). The mesohabitat approach can be considered innovative in the study of bullhead. In fact, only Gosselin et al. (2010) carried out a study on bullhead mesohabitat use in which the fish displayed a strong association with glides (i.e., relatively deep habitats having high rates of velocity increase with flow). Other studies analyzed bullhead preferences at the micro-scale (some of them differentiating adults from juveniles) in relation to a number of variables ranging from four (e.g., Carter et al. 2004; Van Liefferinge et al. 2005) to eight (Davey et al. 2005), with depth, velocity and substrate the ones most commonly used.

To model habitat requirements of *Cottus gobio*, the aims of this study were (i) to identify the most important mesohabitat attributes for bullhead presence and abundance in Alpine streams, (ii) to evaluate the performances of RF and LR in building mesohabitat suitability models and (iii) to investigate possible RF applications in the framework of MesoHABSIM to predict habitat availability in regulated Alpine rivers.

Methods

Study area

The study domain is the mountainous areas of Piemonte and Valle d'Aosta regions (North-Western Italy), which together represent 28 % of the total length of the Alpine mountain range. Alpine watersheds are characterized by low flows in winter affected by snow cover accumulation, presence of glaciers and freezing processes (Vezza et al. 2010). Where snowpack is deep and glaciers are present (Northern Piemonte and Valle d'Aosta), high flows occur in summer mainly driven by ice and snowmelt. In the Southern part of the region (Mediterranean Alps), the river regime can have two peaks: the first high flow usually occurs late in Spring, in May or even June, and the second (which is usually lower) in November (Regione Piemonte 2007). Land use is mainly characterized by rocky and forested areas in the upper part of the catchments, while crops and vineyards constitute the land cover for the lower stretches in river valleys.

Ten watercourses with no or negligible degrees of impact (Fig. 1) were chosen as the environmental reference condition for the development of habitat suitability models according to (i) the natural condition of the flow regime; (ii) the presence of an age-structured fish population; and (iii) the spatial distribution across the region (Vezza et al. 2012a). The study sites were identified on the basis of the three above criteria by analyzing institutional databases (census of water abstractions and river infrastructures, institutional fish monitoring stations, etc.), and consequently by carrying out direct on-site surveys to verify the status of bullhead populations and the absence of infrastructures. Sites were selected only where direct surveys confirmed the presence of the abovementioned environmental reference conditions. This selection followed the virtual reference river concept (Parasiewicz et al. 2008), using the biological needs of desired fauna (e.g. the target fish community) in order to set the foundation of habitat assessment at broad scales. To evaluate the RF applicability in the framework of the MesoHABSIM model, three regulated sites were chosen for models' application (Fig. 1). In these watercourses, bullhead populations are currently almost absent, although expected, due to significant habitat alteration and water withdrawal. Table 1 reports the main hydro-morphological features of the selected streams (i.e. reach elevation, mean channel width and gradient).

Habitat description and fish data

Within representative stream reaches, 95 hydromorphological units (HMUs) were identified and described following the MesoHABSIM approach (Parasiewicz 2007; Vezza et al. 2012b). Surveys were carried out by mapping each HMU in a GIS environment by means of a rangefinder (Trupulse 360B, Laser Technology, Inc., Centennial, CO, USA), a photographic tripod and a rugged field computer (Nomad TDS, Field Environmental Instruments Inc., Sunnyvale, CA, USA, with GPS positioning). Vezza et al. (2012b) reported the advantages of using this habitat description technique, which is particularly suitable for Alpine high gradient streams. It is based on the use of light equipment and can be performed when satellite coverage is marginal or nonexistent or in zones characterized by hiking difficulties among rocks and often by the presence of snow and ice.

HMU types included pool, glide, run, riffle, ruffle, rapid, step-pool, waterfall, backwater and side-arm. To cover the spatial variability of flow conditions in each HMU, between 7 and 30 point measurements of water depth, mean water column velocity and substrate size were carried out (Vezza et al. 2012a). Seven measurements were empirically chosen as the smallest statistically relevant quantity (Parasiewicz 2007). Cover types were identified visually and consisted of 7 categories: boulders, canopy shading, woody debris, overhanging vegetation, submerged vegetation, shallow margins and undercut banks. Physical attributes with many categories (i.e., HMU types and cover) were broken down into multiple variables in binary (No/Yes) format and measurements of depth, velocity and substrate were divided into frequency categories. Lastly, mean channel width, water temperature and reach elevation were included as model inputs to evaluate possible site-scale effects on bullhead distribution. The total list of the collected habitat attributes is reported in Table 2.

Fish data were collected by sampling every HMU with backpack electrofishing (i.e., two-pass removal method, Meador et al. 2003). To assure the direct association between sampled areas and sampled fish species, HMUs were isolated by using nets and, before release within the same sampled HMU, each fish was measured in terms of weight and total length. The total number of captured fish was 166, which were classified into adult and juvenile life stages by means of length/age relationships (Vezza et al. 2012a). Due to the low number of observations of juveniles ($n = 21$), we focused on adult fish (145 individuals).

Data analysis

The association of HMU characteristics with bullhead presence and abundance was explored using multivariate probabilistic models to establish habitat suitability criteria. In particular, Random Forests (RF, Breiman 2001) and Logistic Regressions (LR, Hosmer and Lemeshow 2000) were used to identify habitat attributes influencing the fish

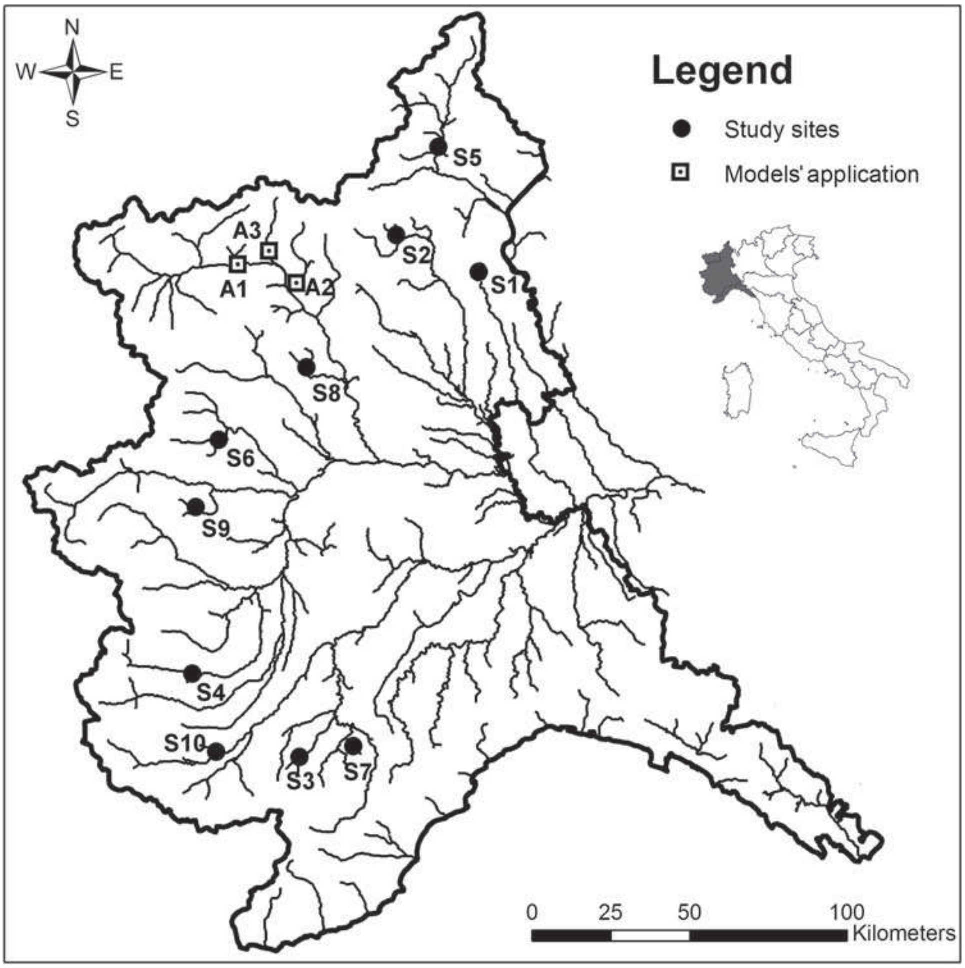

Fig. 1 Study sites for model development and application in NW Italy

Table 1 Study sites selected as environmental reference (S_i) and for models' application in regulated reaches (A_i)

Code	River name	Reach elevation (m a.s.l.)	Mean channel width (m)	Mean channel gradient (%)	Q_{50} (m^3/s)
S1	Agogna	358 ± 0.5	13.2 ± 1.5	2.1 ± 0.6	2.35
S2	Cavaglione	720 ± 2.4	4.5 ± 2.5	8.2 ± 3.1	1.43
S3	Lurisia	634 ± 1.4	6.5 ± 1.3	3.5 ± 1.3	0.87
S4	Melle	666 ± 1.2	5.1 ± 1.3	12.1 ± 2.9	1.32
S5	Ravine	362 ± 2.6	7.3 ± 3.6	10.4 ± 1.7	2.21
S6	Ricchiaglio	628 ± 2.0	5.4 ± 1.7	12.0 ± 2.0	1.24
S7	Rifreddo	442 ± 0.4	4.8 ± 2.4	2.2 ± 0.8	0.29
S8	Savenca	476 ± 1.2	11.2 ± 1.8	3.1 ± 0.9	1.27
S9	Taonere	573 ± 4.5	10.6 ± 2.3	10.1 ± 2.6	1.67
S10	Valle ritta	643 ± 1.7	6.3 ± 1.2	8.5 ± 2.3	1.03
A1	Dora baltea (Nus)	516 ± 0.6	16.2 ± 4.2	1.5 ± 0.4	9.10
A2	Evançon (Isollaz)	662 ± 1.1	8.6 ± 1.8	3.1 ± 1.5	4.08
A3	Marmore (Covalou)	745 ± 4.5	7.1 ± 2.1	5.9 ± 1.9	5.28

Mean values and standard deviations of reach elevation (m a.s.l.), mean channel width (m) and mean channel gradient (%) are shown. Moreover, Q_{50} (m^3/s) are also reported for each site

species distribution. Following Parasiewicz (2007), two different binary models were developed using the data collected during the fish sampling campaigns: an absence/presence model, to distinguish between unsuitable and suitable habitats, and a presence/abundance model, to distinguish between suitable and optimal habitats. The density cutoff value (0.05 individuals/m^2) for low and high abundance was determined as the inflection point of the

Table 2 Habitat variables used for describing the hydromorphologic units (HMUs)

Variable name	Value	Classes	Cathegory/description
HMU type	(No/yes)	10	Pool, glide, run, riffle, ruffle, rapid, step-pool, waterfall, backwater, side arm
HMU Gradient	(%)	1	Water surface gradient of the HMU
Cover	(No/yes)	7	Boulders, canopy shading, woody debris, overhanging vegetation, submerged vegetation, shallow margin, undercut banks
Substrate	(% of random samples)	12	Pelal, psammal, akal, microlithal, mesolithal, macrolithal, megalithal, gigalithal
Water depth	(% of random samples)	9	Classes in 15 cm increments (range 0–120 cm and above)
Flow velocity	(% of random samples)	9	Classes in 15 cm/s increments (range 0–120 cm/s and above)
Mean width	(m)	1	Mean channel width at reach scale
Water temperature	(°C)	1	Water temperature at reach scale
Elevation	(m a.s.l.)	1	Mean reach elevation

See Parasiewicz 2007 and Vezza et al. 2012a for details on variable descriptions

envelope curve of the fish density histograms (Parasiewicz et al. 2007; Vezza et al. 2012a). Although the capture efficiency of adult fish in the two-pass backpack electrofishing ranged from 72 to 100 %, the estimated total fish abundance was considered acceptable to divide bullhead occurrence into three abundance classes. The prevalence (i.e. the frequency of occurrence of the target organism) was therefore 0.42 for the absence/presence model and 0.46 for the presence/abundance model.

Random Forests

As the response variable was a binary variable (fish absence/presence and presence/abundance), we confined our attention to classification RF models. In RF, as implemented in R (library randomForest, version 4.6-7, Liaw and Wiener 2002) each tree is trained by selecting a random bootstrap subset X_i (with i ranging from 1 to t, maximum number of trees) using two-thirds of the original dataset X and a random set of predictive variables. The elements not included in the training dataset are referred to as out-of-bag data (OOB, i.e. cross-validated accuracy estimates) for that bootstrap sample. On average, each element of X is an OOB element in one-third of the t iterations. After growing the forest, global RF accuracies and error rates (i.e. the OOB error, E_{OOB}, and the within-class errors, $E_{Class(j)}$) are finally computed using the OOB predictions (Franklin 2010).

Taking into account OOB error stabilization, the total number of trees (t) was equal to 2000 replicates (Evans and Cushman 2009), whereas the m parameter was defined for each model as the square root of the total number of predictor variables with a minimum of $m = 2$ (Breiman 2001). To assess the importance of a specific predictor variable we used the function “importance” in the randomForest library and, in particular, the metric called Mean Decrease in Accuracy (MDA, Liaw and Wiener 2002). Increase in MDA variable importance indicates the contribution to the RF prediction accuracy for that variable. The most parsimonious model was identified by the Model Improvement Ratio (MIR, Murphy et al. 2010) technique, which uses the MDA variable importance, standardized from zero to one, to define the best parsimonious model. Using this variable ranking as reference, the model variables were subset using 0.02 threshold increments, with all variables above the threshold retained for each model (Evans and Cushman 2009). Each subset model was then compared and the model that exhibited the minimum E_{OOB} and the lowest maximum $E_{Class(j)}$ was selected. Lastly, correlation among selected variables was tested using a correlation matrix to avoid collinearity effects on model performance.

The partial dependence plots (or response curve), based on the RF results, provided a way to visualize the marginal effect of the selected independent variables on the fish distribution (Cutler et al. 2007) outlining the relationships between individual habitat variables and the predicted probabilities of fish presence or abundance.

Logistic regressions

LR have been widely used in the literature for habitat suitability evaluation (Pearce and Ferrier 2000; Filipe et al. 2002; Tirelli et al. 2009). Moreover, model outputs are easily interpretable and are implemented in the MesoHABSIM simulation system (Parasiewicz 2007). In the LR model construction, we first calculated the correlation among potential model inputs by means of a correlation matrix (polycor package, version 0.7-8, Fox 2007). We then used the univariate analysis and the ecological relevance of each parameter derived from previous studies

(Regione Piemonte 2007) to identify, among the correlated variables, the ones that were better predictors of the response variable.

LR models were built using a cross-validation procedure. This approach is frequently used when the number of observations available is not sufficient to separate the data and validate the models using an independent data set (Steyerberg et al. 2001). For each model run, 20 % of the available data was set aside for validation purposes and, with the remaining 80 %, the Akaike's information criteria (AIC, Sakamoto 1994) and a stepwise forward procedure were used to determine which parameters should be included in the model. It is important to note that while the RF ranking of variables was based on all possible combinations of model inputs, the one-step-ahead search procedure of the LR may not lead to the best combination of inputs.

To increase model certainty, the procedure is repeated 20 times and, each time, a new set of randomly selected data is set aside for validation purposes. After 20 runs, the model generates a list of parameters selected in at least two runs and conducts one additional run using only these parameters as input attributes. Standard errors were evaluated to avoid over-fitting and produce parsimonious models (i.e., resulting in a limited number of habitat descriptors) while the Receiver Operating Characteristic (ROC) analysis performed for each predictor was used as a measure of variable importance (function varImp, caret package, version 5.15-61, Kuhn 2008). Finally, the ROC curve and the sensitivity–specificity sum maximization criterion (Liu et al. 2005; Jiménez-Valverde and Lobo 2007) provided the probability thresholds used to assign abundance classes to each mesohabitat observation (fish absence, presence or abundance; Parasiewicz 2007).

Model evaluation

For both RF and LR, the performance of the predictive models was evaluated using five performance metrics (accuracy, sensitivity, specificity, Cohen's kappa and area under ROC curve) based on the results of the cross-validation procedures (Mouton et al. 2010). The estimated accuracy represents the proportion of overall correctly classified observations, while sensitivity and specificity, respectively, refer to the proportion of actual positives and negatives correctly identified as such. The Cohen's kappa coefficient is a statistical measure of inter-rater agreement for categorical items and it is generally thought to be a more robust measure than simple percent agreement calculation since kappa takes into account the agreement occurring by chance. The area under ROC curve (AUC), measured from ROC plots, is a performance metric that is independent of prevalence (Mouton et al. 2010) and represents a useful measure of how well a model is parameterized and calibrated (Manel et al. 2001).

Model application in regulated sites

A comparison of the obtained models was also carried out by means of habitat evaluation in three regulated sites, i.e., Dora Baltea (Nus), Evançon (Isollaz) and Marmore (Covalou, Table 1). Since LR is already implemented in the MesoHABSIM model (Parasiewicz 2007; Parasiewicz et al. 2012b), this analysis was aimed at evaluating possible applications of RF to predict habitat availability in regulated Alpine rivers.

For the three regulated sites, bullhead are considered a present species but local populations are heavily affected by habitat alteration and water abstraction and only a few specimens are occasionally sampled (Regione Valle d'Aosta 2008). The obtained mesohabitat suitability criteria were therefore applied to predict fish distribution (fish absence, presence and abundance) and, consequently, to classify habitat into suitability categories (i.e., not suitable, suitable or optimal habitat, Parasiewicz 2007).

Results

Bullhead were found in 5 of the 10 considered HMU types (i.e. pool, riffle, rapid, ruffle, run). Table 3 describes the main HMU features, including the range of mean depth and mean flow velocity, the dominant substrate and the bullhead presence proportions.

In Fig. 2, the performance of RF and LR for absence/presence and presence/abundance models is compared, highlighting the mean values and one standard deviation for each performance metric. With regards to model performance, our results show that RF outperformed LR for both absence/presence (RF: 84 % accuracy, $k = 0.58$ and $AUC = 0.88$; LR: 78 % accuracy, $k = 0.54$ and $AUC = 0.85$) and presence/abundance models (RF: 79 % accuracy, $k = 0.57$ and $AUC = 0.87$; LR: 69 % accuracy, $k = 0.43$ and $AUC = 0.81$). Moreover, for each model, the confusion matrix was plotted as bar charts (Fig. 3) and the relative importance of variables was standardized to sum to one to compare the relative importance of predictors within a predictor set and across the two statistical methods (Fig. 4).

The mesohabitat suitability models are reported in Fig. 5 for RF and in Table 4 for LR, in which the selected habitat variables are listed in order of importance. MESOLITHAL (6–20 cm) and MACROLITHAL (20–40 cm) substrate types, which were selected by both statistical techniques, demonstrated to be the most important habitat attributes, with a positive influence on bullhead probability

Table 3 Description of the five mesohabitat types in which bullhead were present

Mesohabitat type	Proportion over the total sampled mesohabitats	Range of mean depth (cm)	Range of mean velocity (cm/s)	Dominant substrate (−)	Substrate classes (cm)	Bullhead presence (%)
Pool	23	42–110	0–32	Micro-Meso-Macrolithal	2–6; 6–20; 20–40	22
Riffle	23	18–76	15–75	Meso-Macrolithal	6–20; 20–40	43
Rapid	24	12–60	28–95	Meso-Macrolithal	6–20; 20–40	50
Ruffle	10	13–47	15–48	Meso-Macrolithal	6–20; 20–40	60
Run	11	32–67	16–45	Micro-Meso-Macrolithal	2–6; 6–20; 20–40	64

Bullhead were absent in glides, plunge pools, waterfalls, backwaters and side arms

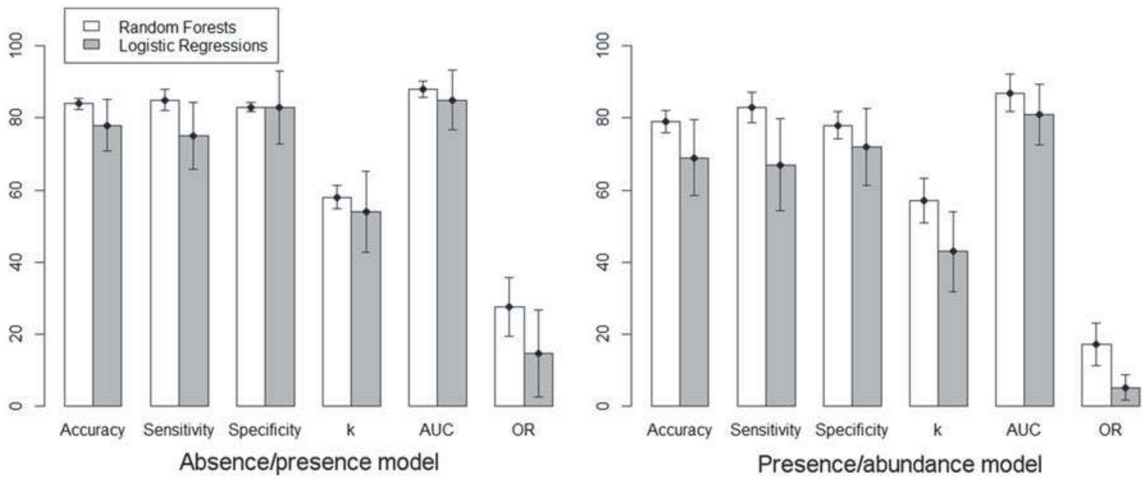

Fig. 2 Random Forests and Logistic Regressions performance for absence/presence and presence/abundance models. *Confidence bars* show the mean values and one standard deviation of cross-validation results in terms of model accuracy (correctly classified observations), sensitivity, specificity, Cohen's kappa (k) and area under ROC curve (AUC)

of presence. The other important attributes defined by RF were HMU gradient, water depth of 60–75 cm and the cover provided by canopy shading, while LR selected flow velocity of 0–15 cm/s, water depth of 15–30 cm and habitat type RUN as significant variables for bullhead presence. For the prediction of fish abundance, both statistical techniques provided almost the same results in terms of selected variables; namely HMU gradient, flow velocity of 0–15 cm/s and water depth of 15–30 cm. In addition, RF abundance model included flow velocity of 30–45 cm/s and LR selected the habitat type RUFFLE (i.e. dewatered rapid) as significant model input.

The models' application in two regulated stream reaches is reported in Fig. 6 (i.e., Dora Baltea—Nus and Evançon—Isollaz) showing the spatial distribution of mesohabitats and the relative model predictions in terms of unsuitable, suitable and optimal habitat. The overall comparison of models' application (three regulated sites) is reported in Fig. 7 by means of a coincidence matrix (Stehman 1997) displayed using bar-charts. The overall agreement between the two statistical techniques in the prediction of mesohabitat suitability was 77 % with a Cohen's kappa equal to 0.59.

Discussion

The present research gave insight into the mesohabitat preferences of bullhead in Alpine streams and compared the accuracies of Random Forests (RF) and Logistic Regressions (LR) to develop habitat suitability models for fish. RF is currently considered a promising technique in ecology (Cutler et al. 2007; Franklin 2010; Drew et al. 2011; Cheng et al. 2012), and although RF has already been applied in freshwater fish studies (Buisson et al. 2010; Grenouillet et al. 2011; Mouton et al. 2011; Markovic et al. 2012), this paper contributed to test the applicability of this statistical technique in the field of habitat-hydraulic modeling. On the other hand, LR has been widely tested and applied on various studies on freshwater fish (Pearce and Ferrier 2000; Filipe et al. 2002; Tirelli et al. 2009) and in the MesoHABSIM simulation system (Parasiewicz 2007;

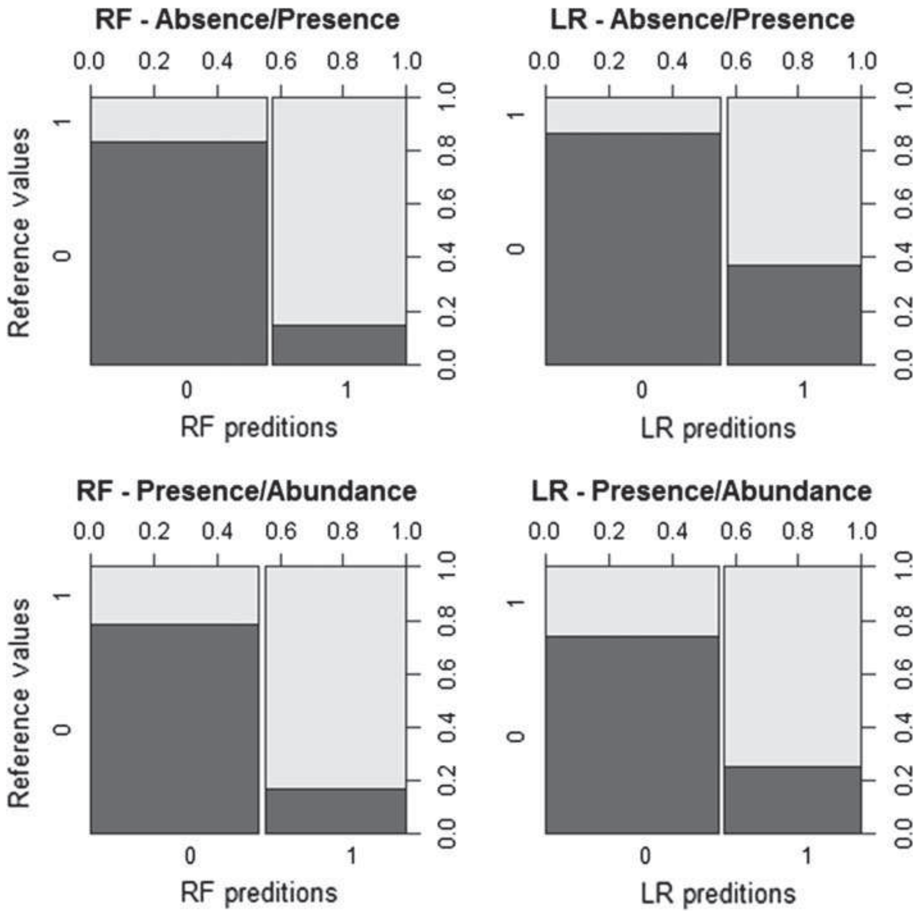

Fig. 3 Confusion matrices of the different models based on Random Forests (RF) and Logistic regressions (LR) modeling techniques. All matrices are reported as bar charts and cross-print reference values and model predictions

Parasiewicz et al. 2012a). We selected the RF technique since it has been successfully applied in ecology and is categorised among the best techniques available for ecological modeling. However, other comparative studies can be performed to test the implementation of other techniques in the frame of habitat-hydraulic models (e.g., Olden and Jackson 2002; Kampichler et al. 2010; Melcher et al. 2012).

Fish data collection and the description of mesohabitat characteristics was carried out in 10 reference streams of NW Italy (Alps mountain range) in accordance to the MesoHABSIM approach (Vezza et al. 2012a). Bullhead were found as present in five types of hydromorphological units (HMUs): pool, riffles, rapids, ruffles and runs, and, in contrast with Gosselin et al. (2010), no bullhead were observed in glides. This result is not surprising due to the important morphological differences between Alpine streams and the Lowland Dowles Brook (Worcestershire, UK, Gosselin et al. 2010) and the limited occurrence of glides in the mountainous area of the Alps (4 % over the total number of sampled HMUs). On the other hand, the presence of bullhead in pools is in agreement with Gosselin et al. (2010) and, furthermore, the higher presence in riffles is supported by previous studies (Roussel and Bardonnet 1996; Langford and Hawkins 1997; Cowx and Harvey 2003; Tomlinson and Perrow 2003) that indicated a marked preference for this HMU type. However, the highest bullhead presence (64 %) was recorded in runs, which were selected by LR as significant for an high probability of bullhead presence.

We evaluated and compared RF and LR based on model performance and the ecological relevance of selected variables. The performance of RF exceeds that of LR in accordance with other comparative studies (Cutler et al. 2007; Siroky 2009) for both absence/presence and presence/abundance model (Figs. 2 and 3). This result underlines the advantages of RF compared to LR in terms of an higher classification accuracy. RF also showed more stable results (lower standard deviation) compared to the LR algorithm (Fig. 2) and should be therefore favored to model fish-habitat relationships because this technique should provide superior results when the data relationships are non-linear (see e.g., the partial plots reported in Fig. 5). As highlighted by Olden and Jackson (2002), patterns within ecological data are commonly non-linear in nature and the non-linear approaches should perform as well as linear methods when the data show linear relationships. It is important to highlight that to overcome the problem of

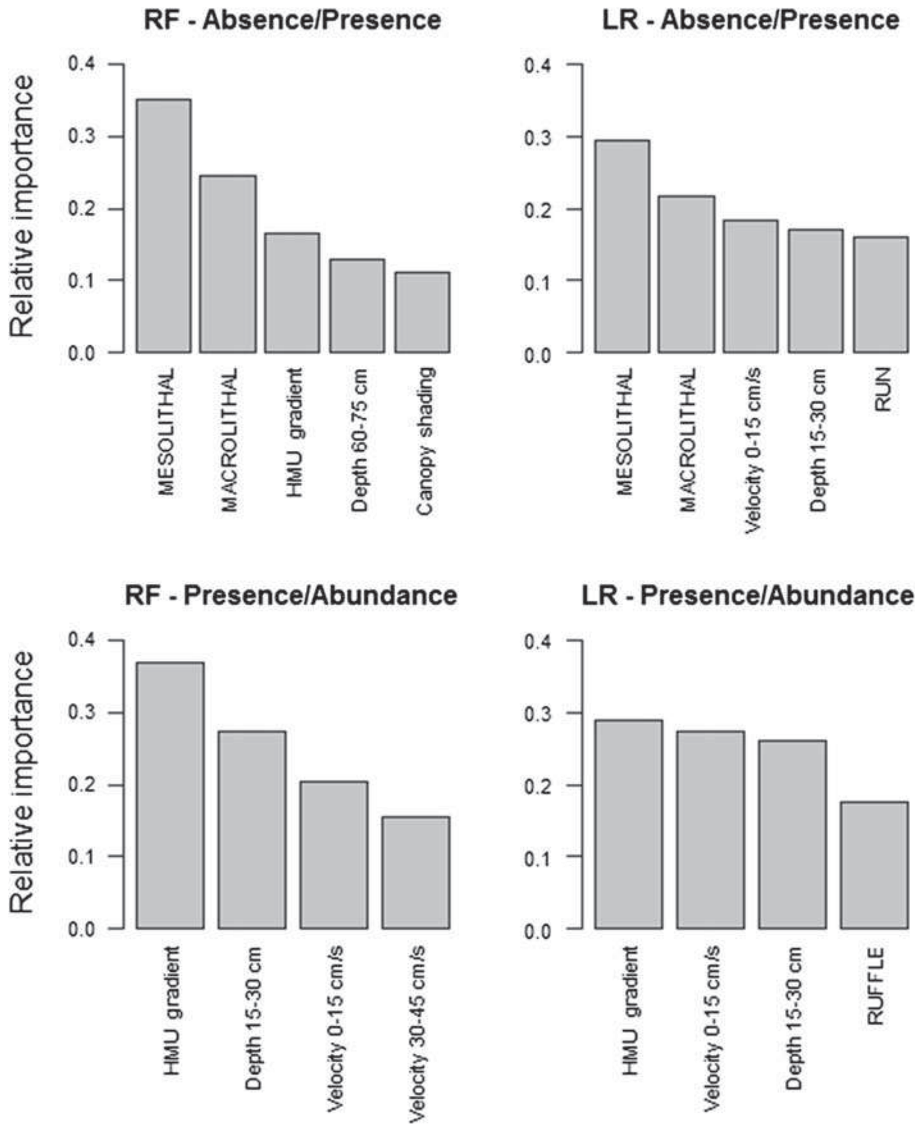

Fig. 4 Relative importance (standardized to sum to one) of each input variable to predict absence/presence or presence/abundance of bullhead in Alpine streams for Random Forests (RF) and Logistic Regressions (LR) models

variability in predictions among different modeling techniques, consensus methods can be used to combine ensembles of species range forecasts and to reduce the uncertainty of results (Marmion et al. 2009). The implementation of these ensemble forecasting of species distributions can be seen as a further development in the field of habitat-hydraulic modeling.

The variable selection procedure was based on two different approaches for the two statistical techniques and led to two different sets of model inputs. This result was confirmed in other studies (Xu and Zhang 2001; Abrahamsson et al. 2003; Reunanen 2003; Wells et al. 2011), in which different variable selection procedures produced similar subsets of variables. For example, MESOLITHAL and MACROLITHAL were the most important predictors in both absence/presence models (Fig. 4). In addition, LR selected velocity of 0–15 cm/s, water depth of 15–30 cm and habitat type RUN as significant variables and the relative importance of these three variables was higher compared to the three remaining predictors selected in the RF model. These substantial differences are also not surprising given the inherent differences between the two types of statistical approaches. The random recursive-partitioning algorithm of RF has a number of advantages over the training algorithms of LR, including its ability to capture non-linear relationships and model complex interactions among predictor variables (Cutler et al. 2007). Therefore, these advantages are most likely responsible for the differences in terms of selected variables and relative importance. It is interesting to note that, during the model construction phase, RF was faster to train and run than LR, having the cross-validation and the variable ranking procedures embedded in the algorithm (R package, random Forest, Liaw and Wiener 2002).

Considering bullhead preference for the three most commonly analyzed variables (depth, velocity and

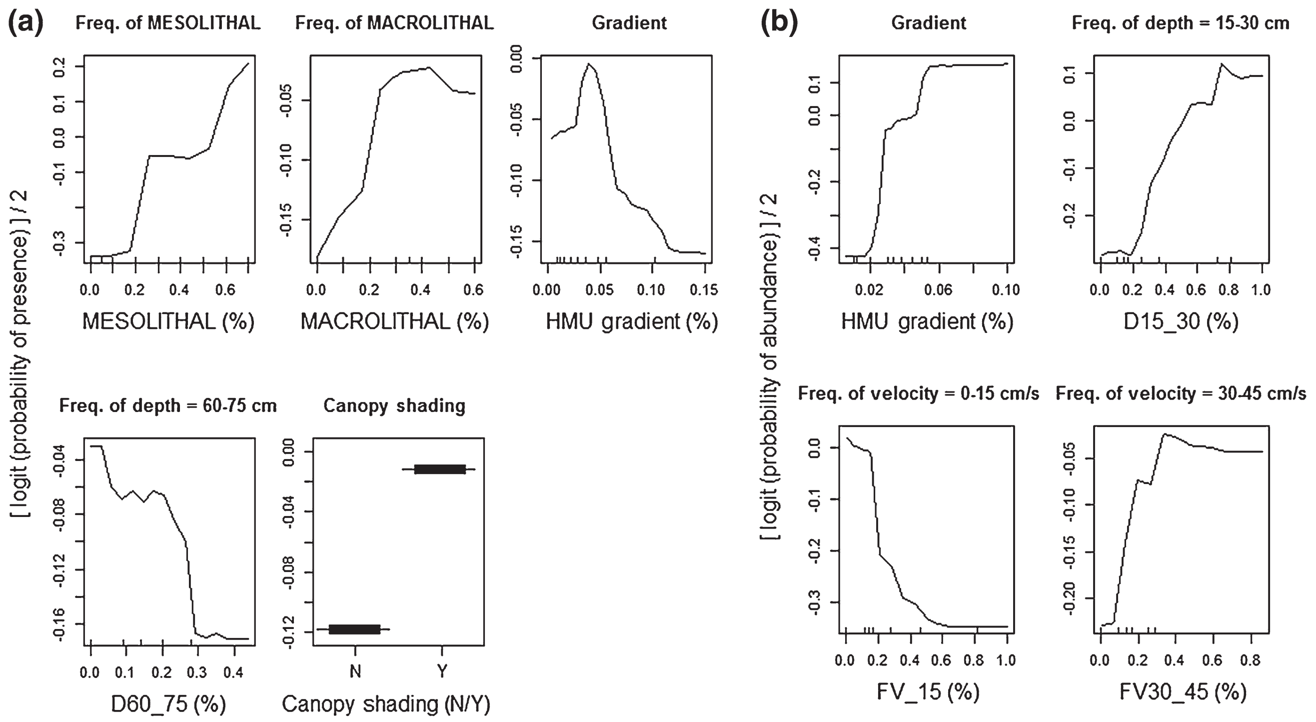

Fig. 5 Bullhead Random Forest model. Selected variables (in order of importance) for (**a**) absence/presence and (**b**) presence/abundance models. The relationship between variables and probability is reported using partial dependence plots to investigate the marginal effect of the selected independent variable on the predicted probability of bullhead presence and abundance (details in, Cutler et al. 2007)

Table 4 Logistic regression models to predict absence/presence and presence/abundance of bullhead

Absence/presence		Presence/abundance	
Probability cutoff	0.46	Probability cutoff	0.54
Constant	−7.06	Constant	−4.21
MESOLITHAL (6–20 cm)	10.00	HMU gradient (%)	7.17
MACROLITHAL (20–40 cm)	7.99	Freq. of velocity 0–15 cm/s	−6.32
Freq. of velocity 0–15 cm/s	−2.24	Freq. of depth 15–30 cm	5.62
Freq. of depth 15–30 cm	−1.84	Ruffle (No/yes)	−3.22
RUN (No/yes)	1.58		

For both models, selected habitat variables are reported in order of importance. The probability cutoff to distinguish between absence/presence and presence/abundance was derived from the Relative Operating Characteristic (ROC) curves, whereas the variable coefficients are multipliers of the significant habitat attributes

substrate), a general agreement among existing studies is only related to substrate, which seems to be the most relevant parameter predicting fish presence: bullhead prefer non-cohesive substrates associated with a wide range of coarse mineral particles like gravel, pebbles, cobbles and boulders (Davey et al. 2005; Legalle et al. 2005a, b; Van Liefferinge et al. 2005; Knaepkens et al. 2006; Gosselin et al. 2010) with adults occupying zones with coarser substrate than juveniles (Davey et al. 2005; Van Liefferinge et al. 2005). Our results further support the above findings, showing that two types of substrates (i.e., MESOLITHAL and MACROLITHAL) were the most important habitat variables in predicting species distribution (Fig. 4), having a positive influence on fish probability of presence in both RF and LR models (i.e., the increasing trend in the RF partial plots, Fig. 5 and the positive variable coefficients in the LR model, Table 4). As a general rule to ensure that bullhead-rearing habitats are available under regulated flow conditions, the presence of MESOLITHAL and MACROLITHAL should be maintained as well as substrate embeddedness should not increase above the natural levels observed.

Flow velocities ranging from 0 to 0.15 m/s were identified by LR models to have a negative influence on both

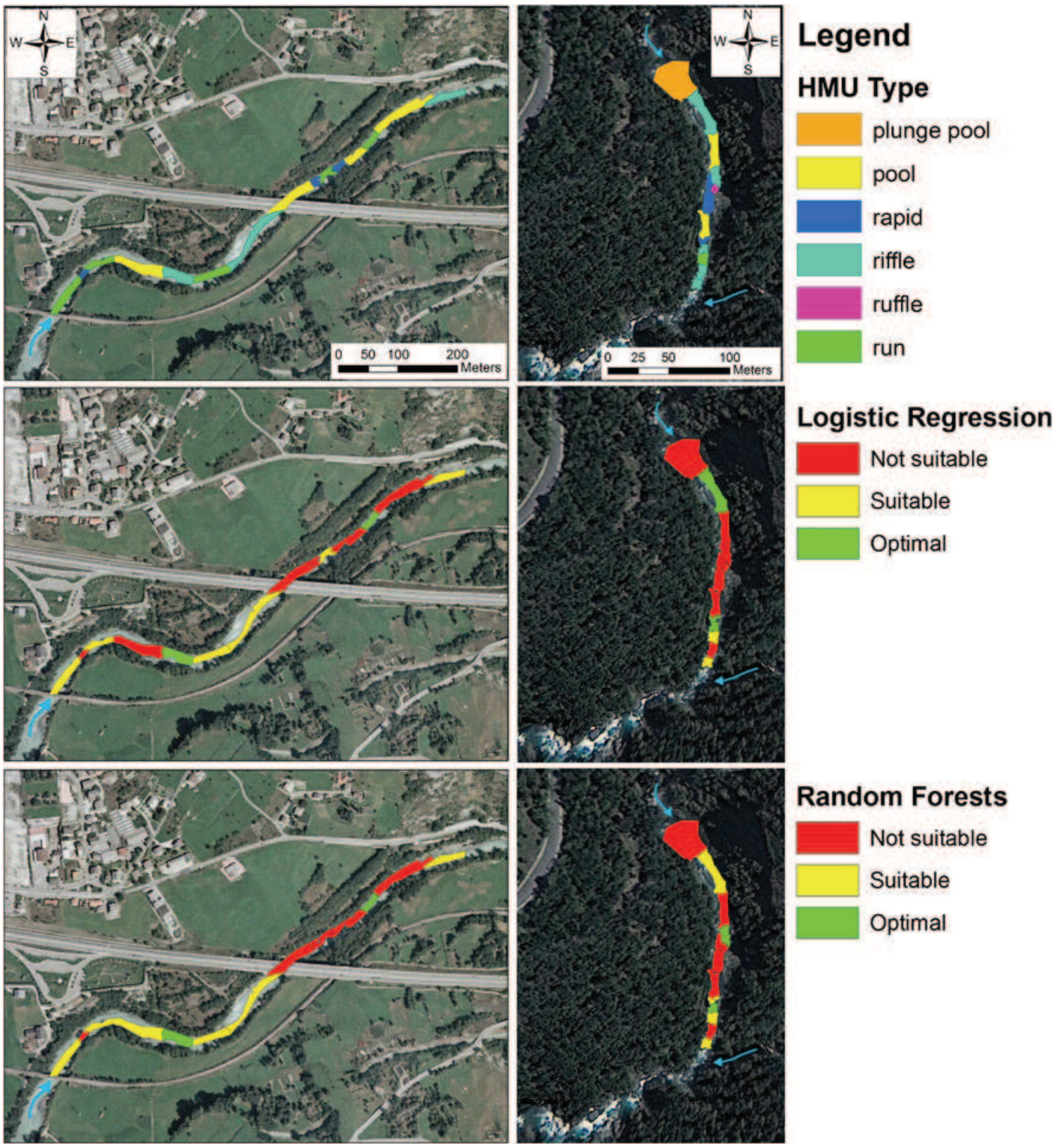

Fig. 6 Model application in two regulated rivers: Dora Baltea (Nus, *left side*) and Evançon (Isollaz, *right side*). Hydromorphological units (HMUs) distribution and model predictions are reported showing the mesohabitat classification into three habitat suitability categories (not suitable, suitable and optimal habitat)

adult bullhead presence and abundance. RF selected the variable flow velocity only for presence/abundance models; in accordance with LR, the velocity range from 0 to 0.15 m/s was identified as significant with a decreasing trend, while flow velocities from 0.30 to 0.45 m/s had a positive influence on bullhead abundance. These results, aligned with the findings of Legalle et al. (2005b) and Knaepkens et al. (2002), are in contrast with Gosselin et al. (2010) and Davey et al. (2005) that indicated a preference for lower values (from 0 to 0.20 m/s): this discrepancy can be attributed to the fact that Alpine streams have significantly different morphological features compared to the two low gradient streams investigated in the cited studies.

Both LR and RF identified depth range from 0.15 to 0.30 m to have a positive influence on bullhead abundance. Moreover, RF results highlighted that depths ranging from 0.60 to 0.75 m have a negative effect on the probability of bullhead presence. Our results are in accordance with previous studies highlighting depth preferences for values varying from 0.05 to 0.40 m and in particular with Legalle et al. (2005a) that identified a clear preference for depths

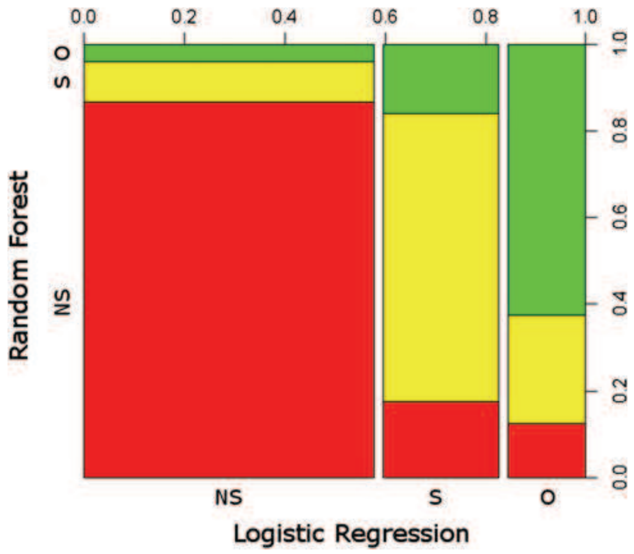

Fig. 7 The comparison of Random Forests (RF) and Logistic Regressions (LR) predictions in three regulated sites displayed by a *bar-chart* coincidence matrix. The *bar* widths refer to the proportions of the predicted habitat suitability categories (i.e., *NS* not suitable, *S* suitable and *O* optimal). The overall agreement between LR and RF was 77 %, with a Cohen's kappa equal to 0.59

ranging from 0.15 to 0.30 m. Also the parameter HMU gradient, as a surrogate of flow velocity, was selected by both LR and RF as having a positive influence on bullhead abundance. Additionally, RF considers it significant also for bullhead presence, with the highest relevance for values around 4 % gradient, which may indicate a physical threshold value. Note that the parameters MESOLITHAL and MACROLITHAL, which demonstrated to be the most important variables for bullhead presence, were not selected in the bullhead abundance model. This can be related to the limited variability of these substrate proportions in mesohabitats where bullhead occurred. Moreover, bullhead were mainly present in mesohabitats characterized by a gradient ranging from 1 to 6 %, and this finding can explain the different trends of bullhead probability of presence and abundance related to the HMU gradient variable (Fig. 5).

Finally, the obtained regional habitat models (both for RF and LR) did not capture the influence of site-scale variables (i.e. channel width, water temperature, reach elevation), which did not seem to be main important habitat characteristics in the analyzed data set. Particularly for water temperature, the exclusion of this variable from the regional habitat models can be related to the limited range of values recorded during the fish sampling campaigns (7.1–15.4 °C).

We also evaluated the models application in three regulated stream reaches to test RF potentials in the framework of MesoHABSIM. Indeed, the application of habitat suitability models to altered instream conditions can quantify habitat availability and can be used for environmental flow assessments (Vezza et al. 2012a) and habitat restoration measures (Parasiewicz et al. 2012b). Although these habitat suitability models make reliable predictions, the applicability of criteria to regulated streams needs to be evaluated and the user may verify the range of conditions over which there is a desire to draw inference from the model. Indeed, such application domain may differ from the original training data in a range of ways and representative test data can be locally required (Vaughan and Ormerod 2005). A model transferability test can be performed if (i) the preferred habitat conditions are available; (ii) the target species abundance and sample sizes are sufficient; and (iii) the influence of biotic factors not described by models (such as competition and predation) preventing animals from using preferred habitats is limited (Thomas and Bovee 1993; Randin et al. 2006). To test the reliability of the obtained habitat suitability criteria across the Alpine area, models' validation will be investigated in other regions of Northern Italy using the proposed approach and surveying techniques as a benchmark for further research studies.

For future applications, it is important to note that RF does not leave polynomial formulas to apply, as in the case of LR, and new data need to pass down through the entire forest to be predicted. Due to this lack of transparency (like for other machine learning techniques, Olden and Jackson 2002), RF can be seen as a black box (Hooten 2011). However, the possibility of using partial dependence plots to investigate the marginal effect of selected variables on fish probability of presence (or abundance, Fig. 5) can be seen as a valuable method to both visualize RF results and to interpret their ecological meaning (Cutler et al. 2007).

The models application to regulated streams demonstrated a relatively high agreement between the two statistical techniques (i.e., 77 % with a Cohen's kappa = 0.59) in habitat suitability classification. Moreover, model results had similar patterns, being predictions of habitat suitability also similar in space. Although RF outperformed LR, the results of this analysis can also underline the potential of LR in building mesohabitat suitability models for fish. Looking at the habitat maps reported in Fig. 6 and the bar-chart coincidence matrix in Fig. 7, one notes the limited amount of suitable and optimal habitats for bullhead in the analyzed regulated reaches. This kind of result can be used therefore to define quantitative benchmarks, goals and targets to guide restoration actions and can contribute to a high potential for designing and monitoring river restoration projects.

The meso-scale resolution allows one to look at the riverscapes of Alpine watercourses as a continuous mosaic of fish habitats, which can be described by a large set of environmental variables. Mesohabitat suitability models are therefore well suited as a planning tool in such environments. Specifically, they can be used for selecting ecologically effective restoration measures and for establishing ecological flow criteria at hydropower or water withdrawals (Vezza et al. 2013). The meso-scale approach demonstrated its particular potential in modelling habitat for fish in Alpine streams and the presented statistical techniques can be considered promising tools for stream ecology management in Northern Italy.

Acknowledgments The presented research was developed in the framework of the HolRiverMed project (275577—FP7-PEOPLE-2010-IEF, Marie Curie Actions) and in collaboration with the Regional Consortium for Fisheries Management in the Aosta Valley (Consorzio Regionale Tutela Pesca - Valle d'Aosta). The data collection was funded by the Regione Piemonte (C61 Project—CIPE 2004) and through the dams monitoring program of Compagnia Valdostana delle Acque (CVA S.p.a.).

References

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

Acreman MC, Ferguson AJD (2010) Environmental flows and the European water framework directive. Freshwat Biol 55:32–48

Ahmadi-Nedushan B, St-Hilaire A, Bérubé M, Robichaud É, Thiémonge N, Bobée B (2006) A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Res Appl 22:503–523

Bain MB, Knight JG (1996) Classifying stream habitat using fish community analysis. In: Leclerc MCH, Valentin S, Boudreau A, Cote Z (eds) Ecohydraulics 2000. 2nd International Symposium on Habitat Hydraulics, INRS-Eau, Quebec City, pp 107–117

Borsányi P, Alfredsen K, Harby A, Ugedal O, Borsányi CKL (2003) A meso-scale habitat classification method for production modelling of atlantic salmon in Norway

Breiman L (2001) Random forest. Mach Learn 45:5–32

Breiman L, Friedman JH, Olshen R, Stone CJ (1984) Classification and regression trees. Wadsworth and Brooks, Monterey

Buisson L, Thuiller W, Casajus N, Lek S, Grenouillet G (2010) Uncertainty in ensemble forecasting of species distribution. Global Change Biol 16:1145–1157

Carter MG, Copp GH, Szomlai V (2004) Seasonal abundance and microhabitat use of bullhead *Cottus gobio* and accompanying fish species in the River Avon (Hampshire), and implications for conservation. Aquatic Conserv 14:395–412

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

Ciuffardi L, Bassani I (2005) Segnalazione del successo riproduttivo della Lampreda di mare (*Petromyzon marinus*) in Provincia della Spezia. Biologia ambientale 19:15–16

Comoglio C, Prato EP, Ferri M, Gianaroli M (2007) A case study of river rehabilitation for fish in northern Italy: the Panaro river project. Am J Environ Sci 3:85–92

Cowx IG, Harvey JP (2000) Monitoring the bullhead, Cottus gobio. Conserving Natura (2000) Rivers. English Nature Peterborough, UK

Crivelli AJ (1996) The freshwater fish endemic to the Mediterranean region. An action plan for their conservation. Tour du Valat Publication

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88:2783–2792

Davey AJH, Hawkins SJ, Turner GF, Doncaster CP (2005) Size-dependent microhabitat use and intraspecific competition in *Cottus gobio*. J Fish Biol 67:428–443

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

Evans J, Cushman S (2009) Gradient modeling of conifer species using random forests. Landscape Ecol 24:673–683

Filipe AF, Cowx IG, Collares-pereira MJ (2002) Spatial modelling of freshwater fish in semi-arid river systems: a tool for conservation. River Res Appl 18:123–136

Fox J (2007) Polycor: polychoric and polyserial correlations. R package version 0.7-5, http://CRAN.R-project.org/package=polycor. Accessed 3 April 2010

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

Frissell C, Liss W, Warren C, Hurley M (1986) A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environ Manage 10:199–214

Gandolfi G, Zenurian S, Torricelli P, Marconato A (1991) I pesci delle acque interne italiane—Italian inland water fishes. Ministero dell'Ambiente, Istituto Poligrafico e Zecca dello Stato, Roma

Gosselin M, Petts G, Maddock I (2010) Mesohabitat use by bullhead (*Cottus gobio*). Hydrobiologia 652:299–310

Gosselin MP, Maddock I, Petts G (2012) Mesohabitat use by brown trout (Salmo trutta) in a small groundwater-dominated stream. River Res Appl 28:390–401

Grenouillet G, Buisson L, Casajus N, Lek S (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography 34:9–17

Hauer C, Unfer G, Tritthart M, Formann E, Habersack H (2010) Variability of mesohabitat characteristics in riffle-pool reaches: testing an integrative evaluation concept (FGC) for MEM-application. River Res Appl. doi:10.1002/rra.1357

Hayer CA, Wall SS, Berry CR Jr (2008) Evaluation of predicted fish distribution models for rare fish species in South Dakota. N Am J Fish Manage 28:1259–1269

Hooten MB (2011) The state of spatial and spatio-temporal statistical modeling. In: Drew C, Wiersma Y, Huettmann F (eds) Predictive species and habitat modeling in landscape ecology. Springer, New York, pp 29–41

Hosmer DW, Lemeshow S (2000) Applied Logistic Regression, 2nd edn. John Wiley & Sons, Inc., New York

Jewitt GPW, Weeks DC, Heritage GL, Gorgens AHM (2001) Modelling abiotic-biotic links in the rivers of the Kruger National Park Mpumulanga South Africa. In: Acreman M, Jewitt GPW (eds) Hydro-ecology: linking hydrology and aquatic ecology. Proceedings of Workshop HW2 held at Birmingham UK July 1999 IAHS Press, Wallingford, pp 77–90

Jiménez-Valverde A, Lobo JM (2007) Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta Oecol 31:361–369

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

Kemp JL, Harper DM, Crosa GA (1999) Use of 'functional habitats' to link ecology with morphology and hydrology in river rehabilitation. Aquat Conserv: Mar Freshwat Ecosyst 9:159–178

Knaepkens G, Bruyndoncx L, Bervoets L, Eens M (2002) The presence of artificial stones predicts the occurrence of the European bullhead (*Cottus gobio*) in a regulated lowland river in Flanders (Belgium). Ecol Freshwat Fish 11:203–206

Knaepkens G, Baekelandt K, Eens M (2006) Fish pass effectiveness for bullhead (*Cottus gobio*), perch (*Perca fluviatilis*) and roach (*Rutilus rutilus*) in a regulated lowland river. Ecol Freshwat Fish 15:20–29

Kuhn M (2008) Building predictive models in R using the caret package. J Stat Soft. http://www.jstatsoft.org/v28/i05/paper. Accessed 22 November 2008

Langford TE, Hawkins SJ (1997) The distribution and abundance of three fish in relation to timber debris and mesohabitats in a lowland forest stream during autumn and winter. Limnetica 13:93–102

Legalle M, Mastrorillo S, Santoul F, Céréghino R (2005a) Ontogenetic microhabitat shifts in the bullhead, *Cottus gobio L.*, in a fast flowing stream. Internat Rev Hydrobiol 90:310–321

Legalle M, Santoul F, Figuerola J, Mastrorillo S, Céréghino R (2005b) Factors influencing the spatial distribution patterns of the bullhead (*Cottus gobio L., Teleostei Cottidae*): a multi-scale study. Biodivers Conserv 14:1319–1344

Liaw A, Wiener M (2002) Classification and regression by Random forest. R News 2:18–22

Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28:385–393

Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–931

Markovic D, Freyhof J, Wolter C (2012) Where are all the fish: potential of biogeographical maps to project current and future distribution patterns of freshwater species. PLoS ONE 7:e40530

Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. Divers Distrib 15:59–69

Meador M, McIntyre J, Pollock K (2003) Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure. Trans Am Fish Soc 132:39–46

Melcher AH, Lautsch E, Schmutz S (2012) Non-parametric methods—Tree and P-CFA—for the ecological evaluation and assessment of suitable aquatic habitats: a contribution to fish psychology. Psych Test Ass Mod 54:293–306

Mouton AM, Schneider M, Depestele J, Goethals PLM, De Pauw N (2007) Fish habitat modelling as a tool for river management. Ecol Eng 29:305–315

Mouton AM, De Baets B, Goethals PLM (2010) Ecological relevance of performance criteria for species distribution models. Ecol Model 221:1995–2002

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

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

Olden JD, Jackson DA (2002) A comparison of statistical approaches for modelling fish species distributions. Freshwat Biol 47:1976–1995

Olden JD, Jackson DA, Peres-Neto PR (2002) Predictive models of fish species distributions: a note on proper validation and chance predictions. Trans Am Fish Soc 131:329–336

Parasiewicz P (2001) MesoHABSIM: a concept for application of instream flow models in river restoration planning. Fisheries 26:6–13

Parasiewicz P (2007) The MesoHABSIM model revisited. River Res Appl 23:893–903

Parasiewicz P, Rogers J, Legros J, Wirth M (2007) Assessment and restoration of instream habitat of the Eightmile River in Connecticut: Developing a MesoHABSIM model. Report 62, p 65

Parasiewicz P, Nestler J, Poff NL, Goodwin A (2008) Virtual reference river: a model for scientific discovery and reconciliation. In: Alonso MS, Rubio IM (eds) Ecological management: new research. Nova Science Publishers, Inc., New York, pp 189–198

Parasiewicz P, Castelli E, Rogers JN, Plunkett E (2012a) Multiplex modeling of physical habitat for endangered freshwater mussels. Ecol Modell 228:10

Parasiewicz P, Ryan K, Vezza P, Comoglio C, Ballestero T, Rogers JN (2012b) Use of quantitative habitat models for establishing performance metrics in river restoration planning. Ecohydrology. doi:10.1002/eco.1350

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133:225–245

Pini Prato E, Comoglio C, Calles O (2011) A simple management tool for planning the restoration of river longitudinal connectivity at watershed level: priority indices for fish passes. J Appl Ichthyol 27:73–79

Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33:1689–1703

Regione Piemonte (2010) Proposta di piano regionale per la tutela e la conservazione degli ambienti e della fauna acquatica e l'esercizio della pesca. Regione Pimonte, Torino

Regione Piemonte (2007) Piano di Tutela delle Acque–Regional Water Protection Plan, Volume D.C.R. n. 117-10731. Torino, Italia

Regione Valle d'Aosta (2008) Progetto di sperimentazione del DMV del gruppo C.V.A. Caratterizzazione Dora Baltea. Monografia 2008, Aosta, Italia

Reunanen J (2003) Overfitting in making comparisons between variable selection methods. J Mach Learn Res 3:1371–1382

Roussel JM, Bardonnet A (1996) Differences in habitat use by day and night for brown trout (*Salmo trutta*) and sculpin (*Cottus gobio*) in a natural brook: multivariate and multi-scale analyses. Cybium 20:45–53

Sakamoto Y (1994) Categorical data analysis by AIC. In: Bozdogan H (ed) Proceedings of the first US/Japan conference on the frontiers of statistical modeling: an informational approach. Kluwer Academic Publishers, Dordrecht, pp 255–269

Siroky DS (2009) Navigating random forests and related advances in algorithmic modeling. Stat Surveys 3:147–163

Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 62:77–89

Steyerberg EW, Harrell FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 54:774–781

Thomas JA, Bovee KD (1993) Application and testing of a procedure to evaluate transferability of habitat suitability criteria. Regul Rivers: Res Manage 8:285–294

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

Tomlinson ML, Perrow (2003) Ecology of the bullhead. Conserving Natura 2000 Rivers Ecology. English Nature, Peterborough

Van Liefferinge C, Seeuws P, Meire P, Verheyen RF (2005) Microhabitat use and preferences of the endangered *Cottus gobio* in the River Voer, Belgium. J Fish Biol 67:897–909

Vaughan IP, Ormerod SJ (2005) The continuing challenges of testing species distribution models. J Appl Ecol 42:720–730

Vezza P, Comoglio C, Rosso M, Viglione A (2010) Low flows regionalization in north-western Italy. Water Resour Manage 24(14):4049–4074

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

Vezza P, Parasiewicz P, Spairani M, Comoglio C (2012b) Meso-scale habitat modelling in Alpine high gradient streams. In: Mader H, Kraml J (eds) Proceeding of the 9th Eco-hydraulics Symposium ISE 2012. Austria, Vienna

Vezza P, Parasiewicz P, Spairani M, Comoglio C (2013) Habitat modelling in high gradient streams: the meso-scale approach and application. Ecol Appl In press

Wells BJ, Yu C, Koroukian S, Kattan MW (2011) Comparison of variable selection methods for the generation of parsimonious prediction models for use in clinical practice. In: SMDM (ed) Proceedings of the 33rd Annual Meeting of the Society for Medical Decision Making. SMDM, Chicago

Xu L, Zhang W-J (2001) Comparison of different methods for variable selection. Anal Chim Acta 446:475–481

Zerunian S (2002a) Condannati all'estinzione? Biodiversità, biologia, minacce e strategie di conservazione dei Pesci d'acqua dolce indigeni in Italia. Edagricole, Bologna

Zerunian S (2002b) Iconography of Italian inland water Fishes. Tipolitografia FG—Savignano s/P (Mo), Roma

Zerunian S (2007) Problematiche di conservazione dei Pesci d'acqua dolce italiani. Biologia Ambientale 21:49–55