2013

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Comparison of approaches for the development of microhabitat suitability models based on fuzzy logic

Corresponding author: Rafael Muñoz Mas

Director: Francisco Martínez Capel

Table of contents:

Abstract

Brown trout (*Salmo trutta L.*) have been used as an indicator of ecological status. Habitat models assess habitat suitability based on physical conditions such flow velocity or water depth are. There are several methodologies to analyse the suitability and to develop habitat suitability models but, at the microscale, the development of continuous univariate Habitat Suitability Curves (HSCs) is by far the most common approach. Two main methodologies exist in the development of HSCs. The first one considers only the conditions observed at the fish locations (Category II ½ HSCs) whereas the second one considers also the conditions observed in the surrounding area (Category III HSCs) Several authors have suggested that considering each hydraulic variable independently may be questionable. Therefore the use of multivariate approaches among researches have increased. The Fuzzy logic is one of those who has most successfully been applied. The fuzzy logic approach mimics the human reasoning thus are presented in an IF-THEN sequence. If certain conditions are resent then the habitat suitability is that. There are two main approaches in the development of Fuzzy logic models; the Expert-knowledge and the Data-driven. The Expert-knowledge approach is based on the literature and the consensus of scientists whereas the Data-driven approach is based on the optimization of the elements of the model based on field data.

This study presented a methodology to develop Expert-knowledge fuzzy models based on HSCs and compared the results with those derived from the Data-driven approach. Specifically Three habitat suitability models were develop for the three considered size classes; brown trout adult-large (> 20 cm), juvenile-medium (20 - 10 cm) and fry-small (< 10 cm). Two models based on the Expert-knowledge approach but differing on the HSCs, Category II ½ HSCs or Category III HSCs and another model was based on the Data-driven approach. The 9 developed models were spatially explicitly validated in an independent river reach and their performance was compared by means of the fuzzy Kappa statistic.

The Expert-knowledge approach herein presented have demonstrated satisfactory. It showed generally a good performance and did not differed substantially in comparison with the Data-driven approach despite the Expert-knowledge models based on Category II ½ HSCs underrated the deep areas in the adult and juvenile. The Category III based models presented better performances that the Category II ½ counterparts and the models for adult and fry were recommended for further analysis. However the Expert-knowledge models presented lower specificity in comparison with the Data-driven approach. Then, in the juvenile case the Data-driven fuzzy model was de recommended for further analysis. The comparison between models based on the fuzzy Kappa did not showed any similarity and the spatially explicit validation have been demonstrated fundamental in the proper selection between the developed models.

1 Introduction

Freshwater fish are considered to be good indicators of water quality in river systems (Karr, 1981; Angermeier and Davideanu, 2004). Habitat models assess habitat suitability for freshwater fish species based on species selection of physical conditions, e.g. at the microhabitat scale such variables can be flow velocity, water depth or substrate (Bovee, 1982). The suitability of a given variable within the considered range is usually mathematised in an index indicating the degree of suitability of the considered variable; for instance the aforementioned, depth, velocity or substrate, as they take their feasible values. Accordingly to Waters (1976), who firstly introduced the use of suitability curves, these indices assessing the degree of suitability provide an output value that usually ranges from zero and one, with zero being unsuitable and one fully suitable for the target species. There are several methodologies to analyse the suitability and to develop the habitat suitability criteria, but the continuous univariate Habitat Suitability Curves (HSCs) are by far the most common in studies involving the physical habitat simulation (Payne, 2009).

The HSCs are mostly based on the frequency analysis of field data considering the physical properties at the locations where fish were observed and/or the hydraulics at the surrounding unoccupied locations. Fig. 1 shows an example of the typical HSC shape for the variable depth, for the brown trout fry. The optimal depth was determined in the interval around 0.25 m whereas it become straight unsuitable as the depth decreases, but presented a smoother decrement if the depth increases.

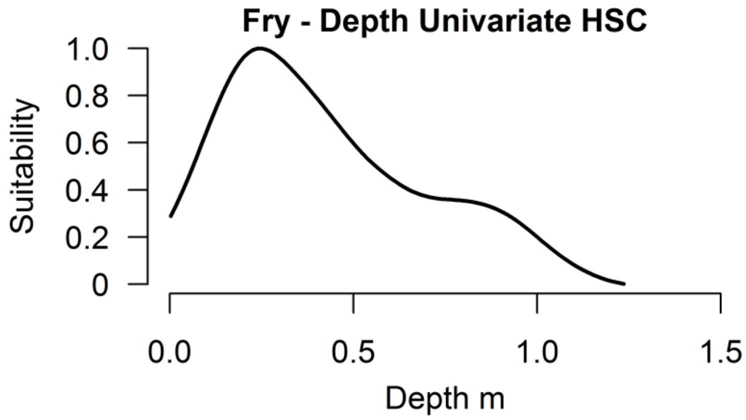

Fig. 1 General example of a univariate Habitat Suitability Curve.

Brown trout have been used as an indicator of ecological status (Karr, 1981) and, for several decades, researchers have developed brown trout habitat models in the form of the aforementioned univariate HSCs (Bovee, 1978; Raleigh, 1984; Heggenes *et al.*, 1991; Rincon and Lobon-Cervia, 1993). Therefore it could be considered a well-known species and the databases related with could be suitable to be analyzed with untested or brand-new statistical methodologies in order to evaluate or improve their performance. Despite the existence of abundant studies, several papers showed difficulties to transfer the developed models, questioning its generalization ability (Mäki-Petäys *et al.*, 2002; Fukuda, 2010). Although there have been reported some succeeds in the development of general models (Nykänen and Huusko, 2004). Multiple factors affect fish habitat selection, thus it is generally recommended the generation of site-specific models of habitat suitability, especially for the application of physical habitat modelling (Moyle and Baltz, 1985; Bovee *et al.*, 1998; Rosenfeld *et al.*, 2005).

The aforementioned HSCs were categorized by Bovee (1986) in three Categories according to the methodology applied to its development. Category I include curves generated from literature and experts consensus. Category II are the Use curves, based on frequency analysis of the hydraulics over the fish locations and does not include any reference to fish electivity or habitat availability; Category III are the preference curves and are derived also from observational data on habitat use (i.e. hydraulics in the locations where fish were observed) but weighted by the habitat availability (i.e. hydraulics over the surrounding unoccupied locations) by calculating the forage ratio (Voos, 1981). Several authors considered an extra category which could be considered a modification of the Category II curves, so called Category II ½. This approach differs from the Category II curves in the way the survey is carried out applying the equal-effort approach (Johnson, 1980) balancing the different combinations of variables in order to avoid any bias derived from the data collection.

The suitability index derived from each of the considered HSC should be summarised in a single index in order to assess the considered area. This composite index, usually called the Habitat Suitability Index (HSI) (sensu Vadas R.L and Orth, 2001), also ranges from zero to one with similar meaning. Different methods have been used to carry out the combination of the different suitability indices obtained from each physical variable in order to produce the HSI. The most important methods to combine them are: the lowest (Korman, 1994), the product (Bovee, 1986), the arithmetic mean (Terrell, 1984) and the geometric mean (Terrell, 1984). The lowest is a 'controlling method' assuming that the most limiting factor determines the upper limit of habitat suitability, then high suitability in any variable cannot compensate the low suitability in other at a given microhabitat. The product method is also a 'controlling method', whereas arithmetic and geometric means are partially 'compensatory methods' (U.S. Fish and Wildlife Service 1981). The product approach assumes that unsuitable habitat conditions based on one variable cannot be balanced by good conditions based on others (Bovee, 1986). In contrast, the arithmetic mean assumes that good habitat conditions based on one variable can compensate for poor conditions of another (Terrell, 1984). Finally, the geometric mean assumes that each environmental variable is equally important (Benaka, 1999; Rubec, 1999) thus unsuitable conditions derived from a given variable could be also compensated by the remaining variables as well as the arithmetic mean does.

Despite the existence of this wide range of possibilities, several authors have suggested that considering each hydraulic variable independently may be questionable so, it could induce a bias as a result of overlooking possible interactions between variables (Orth and Maughan, 1982; Lambert and Hanson, 1989), because fish do not select the habitats based on a single variable, but in a group of environmental variables that they can evaluate and balance. Therefore, researchers have developed and successfully applied multivariate techniques which are able to model fish habitat suitability taking into account the interactions between variables (Hayes and Jowett, 1994; Lamouroux *et al.*, 1998; Vismara *et al.*, 2001; Ayllón *et al.*, 2010; Muñoz-Mas *et al.*, 2012). One of these techniques, the fuzzy logic approach, was demonstrated to be useful in ecological modelling at different scales and life forms. At the mesoscale, fish and also macroinvertebrates habitat suitability have been successfully modelled (Mouton *et al.*, 2009; Mouton *et al.*, 2011), and at the microscale there are also several examples for these organisms (Van Broekhoven *et al.*, 2006; Mouton *et al.*, 2008;

García *et al.*, 2011), even some of them specifically for brown trout (Jorde, 2001; Magdaleno Mas and Martínez Romero, 2005; Muñoz-Mas *et al.*, 2012).

The fuzzy logic approach was firstly developed by Zadeh (1965). An important advantage of the fuzzy logic approach lays on its transparency, which may stimulate communication of model results to stakeholders (Adriaenssens et al., 2004; Van Broekhoven et al., 2006), and its ability to incorporate the ecological gradient theory (Mouton, 2008). The fuzzy logic approach mimics the human reasoning, and it can be communicated in linguistic terms. This approach considers that IF a number of elements exist, THEN a phenomenon occurs. For instance; if velocity is low, depth is medium and the substrate is coarse, then the brown trout will be probably present.

However the fuzzy logic approach is not as imprecise apparently is transforming these descriptions into a mathematical framework (hereafter fuzzy inference system) in which suitable data processing can be performed providing numerical outputs (Kampichler *et al.*, 2000). A fuzzy inference system consists of three parts: (i) fuzzy input and output variables, discretised in Fuzzy Sets (FS) (ii) the Fuzzy Rules and (iii) the fuzzy inference method (Kasabov, 1998). To implement the first part, fuzzy systems categorize the input and the output variable in linguistic terms; the aforementioned: Low, Medium, High etc. defined by Fuzzy Sets (Zadeh, 1965). The second part of the fuzzy inference system is implemented by defining relationships among these categories, by defining rules of association; the Fuzzy Rules (FR). These rules are constructed as the aforementioned IF-THEN sequence, where the 'IF' part is the antecedent and the 'THEN' part is the consequent. The third part consists of the defuzzification procedure giving a single suitability index similar to those obtained for the combination of the partial results from HSCs thus providing directly the HSI.

There are several methodologies to defuzzify, the most commonly applied at this time is the Centre of Gravity (Ahmadi-Nedushan *et al.*, 2008) and has been previously applied in several studies (Jorde, 2001; Muñoz-Mas *et al.*, 2012). The Fuzzy Sets and the Fuzzy Rules of a fuzzy model can be derived based on two main approaches; The Expert-knowledge and the Data-driven, although the border between them could be imprecise and combinations of both have been also suggested (Mouton *et al.*, 2009). Both approaches have been useful in fish habitat modelling (see aforementioned examples) but there is a certain controversy about which one is more accurate.

The Expert-knowledge approach is based on the literature and the consensus of scientists, whereas the Data-driven approach, usually applied in fish habitat modelling, is based on the optimization of the discretization of the input variables developed by Mouton (2008) (i.e. optimization of the number of Fuzzy Sets and their membership functions) and the optimization of the Fuzzy Rules developed by Mouton et al. (2008) (i.e. determination of the proper consequent for each combination of input Fuzzy Sets; the Fuzzy Rules).

The main effort in the development of fuzzy inference systems with the Expert-knowledge approach lays in the proper definition of the Fuzzy Sets and their corresponding membership functions, and in the fact that the number of rules grows exponentially with the number of variables and the amount of Fuzzy Sets, which could result in the redundancy of rules and misinterpretations or wrong formulisations (Chen and Mynett, 2003). On the positive side, the Data-driven approach imply the owning of a large database to optimize the Fuzzy Sets and the Fuzzy Rules, and sometimes it is impossible due to budget limitations leading to short or imperfect databases (Mouton et al., 2008). This is especially important when no training cases appear to assess the proper consequent of one or more rules, which remain undetermined.

Some authors have demonstrated that expert judgment is convergent and Expert-knowledge fuzzy inference systems do not differ substantially depending on the consulted expert (Ahmadi-Nedushan *et al.*, 2008), providing reliable models, while other suggested the opposite (Acreman and Dunbar, 2004; Adriaenssens *et al.*, 2004), supporting the Data-driven approach. In order to shed some light on this kind of discrepancies, some authors pointed out the necessity of evaluating habitat models with independent data to test the reliability of habitat suitability models (Guisan and Zimmermann, 2000), but despite its importance and the existence of some examples, e.g. Guay *et al.* (2000), it could not be considered widespread. Regarding the Expert-knowledge approach, the aforementioned HSCs, usually presented as category II or III suitability curves, could become a general template to develop fuzzy inference systems, because they provide important information for the variable discretization (development of Fuzzy Sets) and their habitat suitability for individual variables as if they were independent. Although preferences curves (Category III) have been strongly criticized (Hayes and Jowett, 1994; Payne, 2009) and its application have been deprecated because of its overcorrection (Bovee, 1996), they present in some cases widely generalization behaviour (Hayes and Jowett, 1994) and in fact presented larger similarity with the Fuzzy Rule optimization procedure programmed in the present study.

2 Objectives

i. The main objective of the present study laid on the comparison of the Expert-knowledge approach and the Data-driven approach in the development of habitat suitability models for three size classes (namely, fry 0-10 cm, juvenile 10-20 cm, adult > 20 cm) of the brown trout (*Salmo trutta L.*).

ii. The Expert-knowledge fuzzy models were based on univariate Habitat Suitability Curves (HSCs) in a sort of quasi Data-driven approach. The HSCs were developed on purpose for the present study. Two Expert-knowledge fuzzy models were developed, the first one based on HSCs with equal effort approach (Category II ½) and the second one based on preference curves (Category III).

iii. Data-driven fuzzy models were also generated for each size class of the brown trout, using the same database used in their respective HSCs set development.

iv. The robustness of the developed Expert-knowledge fuzzy models was tested by training the Fuzzy Rules by means of the Data-driven approach. The developed Fuzzy Rules were compared.

v. Finally, the results were compared among the approaches, and the models were evaluated in a spatially explicitly context, in an independent reach of a similar river; the Cabriel River.

3 Methodology

3.1 Microhabitat data collection

The Mediterranean brown trout (*Salmo trutta L.*) was the target species of the microhabitat study. The fish were previously classified in three size classes, approximately corresponding to small 0-10 cm, medium 10-20 cm, large > 20 cm. However to improve the understandability of the text these size classes will be called, fry, juvenile and adult in accordance with previous studies which classified the brown trout individuals in three size classes (Bovee, 1978; Ayllón *et al.*, 2010). However the present study did not carry out any analysis aimed to identify the relationship between size and age and due to the variability expected in Mediterranean systems certain differences with the aforementioned studies could be expected.

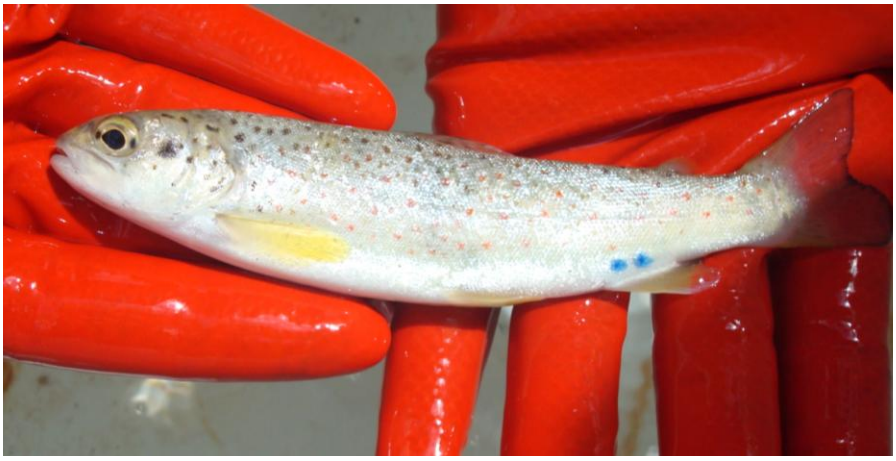

Fig. 2 Juvenile brown trout living in the Cabriel River (showing two marks in Alcian blue, for another kind of study).

The surveys were carried out in two rivers of the Tagus River Basin (TB), Guadiela and Cuervo, and four rivers belonging to the Jucar River Basin District (JRBD); Senia, Turia, Jucar and, its principal tributary, the Cabriel River; Fig. 3.

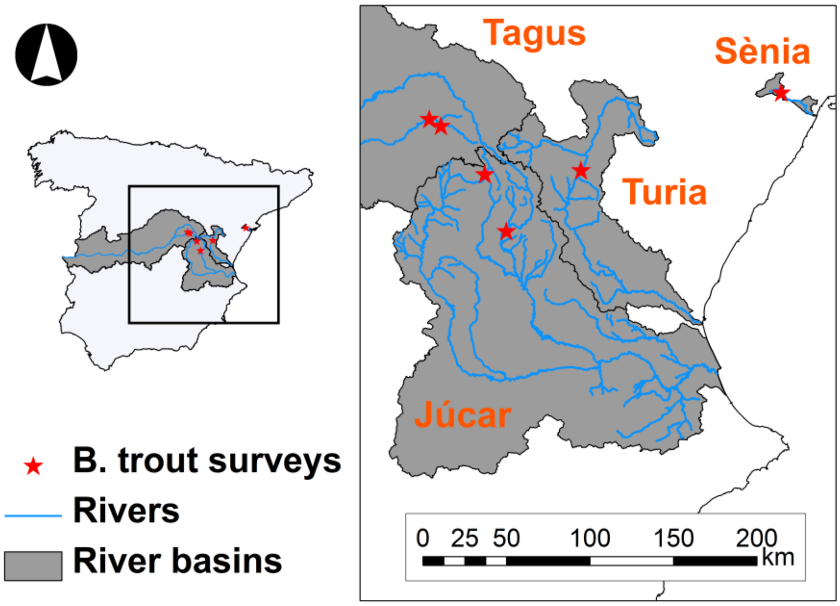

Fig. 3 Map of the sites where brown trout microhabitat surveys were carried out during the period 2005 - 2009 in rivers of the Tagus River Basin (TB) and Jucar River Basin District (JRBD).

The surveys were carried out at low flows during late spring, summer and early autumn in the period 2005-2009. The microhabitat study was done in complete and connected HydroMorphological Units (hereafter HMUs) classified as: pool, glide, riffle, and rapid accordingly to previous studies that classified HMUs in the Mediterranean context (Alcaraz-Hernández et al., 2011). The equal effort approach and the concept of habitat selection (Johnson, 1980) were adapted to the conditions in these Mediterranean rivers, with the selection of equal area (more or less 10 %) of slow and fast water HMUs, grouping pools with glides (slow) and riffles with rapids (fast).

Each HMU was surveyed by underwater observation by snorkelling during daylight, with minimum disturbance to the fish according to standard procedures (Heggenes *et al.*, 1990; Heggenes, 1991). This technique allows the observation of the fish behaviour and its position in the water column, even in relatively adverse surveying conditions (Heggenes *et al.*, 1990; Martínez-Capel and García de Jalón, 1999; Martínez-Capel, 2008). The direct underwater observation has been demonstrated to be more reliable than electrofishing in the location of the fish, there is no displacements because of the galvanotaxis phenomena (Bovee and Cochnauer., 1977; Bovee, 1986; Gatz Jr *et al.*, 1987; Heggenes, 1991), during direct underwater observation it is possible to observe the fish behaviour, then the data are recorded for a specific activity (e.g., holding position and feeding) and no data is recorded if the fish was disturbed. It is very important the experience of the observer, and to spend many hours in the water, performing observation, in order to properly observe the fish activity, elucidating anomalous behaviour derived from the presence of the observer discarding any data related to suspicious or disturbed fish.

Despite we did not perform any transferability test (Thomas and Bovee, 1993), the microhabitat conditions over the entire HMUs were originally measured in cross-sections with a minimum amount of 300 points of unoccupied locations per survey, hereafter Availability records, in order to ensure the applicability of the aforementioned transferability tests thus conditioning the following steps in the development of the fuzzy models (i.e the Data-driven approach needed the application of a sub-sampling methodology). This methodology produced a variable density of data ranging from 1.23 m^2 to 7.96 m^2 per record. Table 1 shows a summary of the sample sizes of each database. All the measurements in the Availability survey and in the Use survey (i.e.

locations where fish were observed) were taken with the same methods and the results were included in the respective datasets, i.e. Availability, no fish observed, or Use, some fish observed.

Table 1 Summary of the sample sizes of microhabitat data for the three size classes, collected over the period 2005 -2009. JRBD: Jucar River Basin District; TB: Tagus Basin.

		Small-Fry (0 - 10 cm)		Medium-Juvenile (10- 20 cm)		Large-Adult (>20 cm)	
River	Period	Use	Availability	Use	Availability	Use	Availability
Jucar(JRBD)	2006 - 2007	19	735	38	735	7	735
Guadiela (TB)	2009					51	455
Turia (JRBD)	2006	25	379				
Cabriel (JRBD)	2005			68	532		
Senia (JRBD)	2006-2007			34	711	11	714
Cuervo (TB)	2009					29	385

Three physical variables were measured to characterise the microhabitats; flow velocity, water depth and substrate type, usually considered the most relevant variables for fish species distribution at this scale in combining with the cover (Waters, 1976; Bovee, 1986; Heggenes, 1990; Gibson, 1993; Bovee, 1998). Although the present study did not included the later due to budget limitations and the inherent difficulty to determine the cover availability based on variations of the current flow. Velocity was measured with an electromagnetic current metre (Valeport®) and depth was measured with a graduated rod at the nearest cm (Fig. 4).

Fig. 4 Current meter and general view of the surveyor carrying out direct underwater observation.

The percentages of each substrate class were visually estimated within 15 cm around the sampling point or at fish location (Bovee and Zuboy, 1988). The classification was simplified from the American Geophysical Union size scale in: bedrock, large boulders (>1024mm), boulders (256–1024 mm), cobbles (64–256 mm), gravel (8–64 mm), fine gravel (2–8 mm), sand (62 mm–2 mm), silt (< 62 mm) and vegetated soil (i.e. substrate covered by macrophytes), similarly to previous works made by snorkelling in Iberian rivers (Martínez-Capel and García de Jalón, 1999; Martínez-Capel *et al.*, 2009b). In order to obtain a single index for mathematical purposes, substrate composition was converted into a single Substrate index (S) by summing weighted percentages of each substrate type. The weights used were: S = 0.08 x bedrock + 0.07 x boulder + 0.06 x cobble + 0.05 x gravel + 0.04 x fine gravel + 0.03 x sand (Mouton *et al.*, 2011).

Table 2 Maximum surveyed velocity and depth and the most abundant substrate per river, derived from all data (Use and Availability).

River	Velocity (m/s)	Depth (m)	Substrate (S)
Cuervo (TB)	1.031	1.24	5
Guadiela (TB)	1.153	1.78	6
Senia (JRBD)	1.755	1.45	7
Jucar (JRBD)	1.284	1.01	0
Cabriel (JRBD)	1.78	1.47	7
Turia (JRBD)	0.621	1.11	0

3.2 Validation data collection

The validation of the performance of the developed models with independent data was based on the assessment of an hydraulic model were trout coordinates were collected for a given flow, thus becoming in an spatially explicit validation.

In previous studies, a 2D hydraulic simulation with River-2D© (University of Alberta 2002) was done in an approximately 300 m long reach of the Cabriel River, 9 km downstream the locations of the microhabitat survey; see more details in Muñoz-Mas *et al.* (2012). The topographic data of the river channel and banks were collected using a Leyca© total station and the substrate composition was visually estimated as aforementioned. Eleven cross-sections at three different flow rates, 0.54, 1.04 and 2.75 m^3/s, were used to calibrate the model in terms of water depth and velocity patterns, accordingly to previous studies (Jowett and Duncan, 2011).

A survey in this river reach was conducted in a single week in the early summer of 2012, with a steady flow rate of 0.89 m^3/s corresponding to the Q_{85} of the stream flow series. Unlike the previous biological surveys, the surveyor did not snorkel all the entire HMUs; instead, the survey was done covering the whole area included in the aforementioned hydraulic model which presents a stable trout population (Martínez-Capel *et al.*, 2009a). The survey was done with similar standard procedures (Heggenes, 1991) as the microhabitat surveys presented in the previous section. Instead of collecting the microhabitat data by recording velocity, depth and substrate, the coordinates in terms of (X,Y,Z) of the observed trout were measured with a FOIF © Total Station; the station was set using the permanent landmarks placed during the topographic survey carried out in the development of the hydraulic model. That (X,Y,Z) information was also used to check the model reliability in terms of morphologic changes on the river bed, showing no lateral displacements of the river channel and a small change in terms of river bed elevations about 0.04 ± 0.12 m, which was considered acceptable; thus the model was considered suitable to be used in the present study.

The information to validate the habitat suitability models by size classes (in terms of velocity, depth and substrate) was, therefore, based on the hydraulic simulation, showing a maximum depth of 1.4 m and 0.53 m/s of mean velocity. The sample sizes for adult, juvenile and fry size classes were 31, 30 and 79 respectively.

3.3 Development of Habitat Suitability Curves

The Habitat Suitability Curves were the input in the generation of the Expert-knowledge fuzzy models in this study.

Two sets of Habitat Suitability Curves of Category II ½ and Category III (Bovee, 1986) were developed, on purpose, for each size class including a curve for velocity, depth and substrate. The procedure to generate curves followed the common standards (Bovee, 1986). The data from each study site were weighted by the surveyed area, in order to equal the degree of influence of each river on the resulting curve. This was followed by a frequency analysis of each separate variable producing the corresponding histogram. The intervals used in the frequency analysis were 5 cm/s for velocity and 5 cm for depth and the intervals for substrate were the nine classes corresponding to the integer numbers of the substrate index. This frequency analysis was carried out for the Use data and the Availability data. The resultant histogram for both datasets were standardised between zero and one. That histogram on the Use dataset was the used to develop the Category II ½ curves.

The Category III curves were developed using the forage ratio (Savage, 1931; Cock, 1978). Some authors (Hayes and Jowett, 1994; Bovee, 1998; Payne, 2009) pointed out the forage ratio may distort the results in terms of suitability (due to over-correction), if either the used or available habitat is poorly represented over any part of their range. In this study, the extreme values of the data distribution (Use and Availability data) were not trimmed for the development of the Category III curves, because the application of the generated Category III curves (in physical habitat simulation) was not an objective of the present study.

For both sets of curves, the Category II ½ and the Category III, a smoothing technique in R environment was applied, specifically the smooth.spline function in the stats package (R Development Core Team, 2012). This procedure was applied to get a unimodal curve, reducing the effect of data gaps in the histogram, thus eliminating some steep segments up and down without ecological sense. In addition, these smoothed curves match the format required for the development of the Expert-knowledge fuzzy models using the herein presented methodology.

3.4 Expert-knowledge fuzzy approach for habitat suitability modelling

The Expert-knowledge fuzzy models were based on the aforementioned univariate Habitat Suitability Curves (HSCs) and the expert judgment of the corresponding authors. The Expert-knowledge fuzzy models present the usual elements in a fuzzy inference system. Firstly, the input variables in the form of categories defined in terms of Fuzzy Sets and their Membership Functions (MF) (Zadeh, 1965). Each Fuzzy Set is mathematically described by their Membership Function which indicates the membership degree, ranging from zero to one, to each Fuzzy Set of a given variable value. Since membership functions have overlapping boundaries, a given value may belong, with different proportions, to two adjacent Fuzzy Sets. Secondly, the set of rules, the Fuzzy Rules, relating each combination of the categories of the input variables with the corresponding output. Both elements present its own development methodology.

The selected geometry of the MF was the trapezoidal, which showed successful in previous studies (Van Broekhoven *et al.*, 2006; Mouton *et al.*, 2007; Mouton *et al.*, 2008; García *et al.*, 2011). A trapezoidal Fuzzy Set is defined by four parameters; a_m, b_m, c_m and d_m. The membership degree to a given Fuzzy Set increases from zero to one between a_m and b_m, is equal to one from b_m to c_m and decrease from one to zero from c_m to d_m. The region between a_m and b_m is shared with the adjacent, if exists, becoming a fuzzy region. The region between b_m and c_m belongs, 'purely', to that Fuzzy Set whereas the region between c_m and d_m is again shared with the adjacent, becoming a fuzzy region as well (Fig. 5).

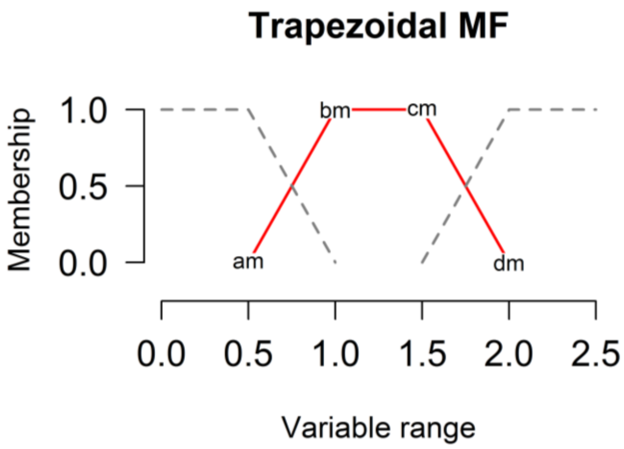

Fig. 5 Example of trapezoidal Membership Function (MF) to a given Fuzzy Set. The region between a_m to b_m and c_m to d_m is shared with the adjacent whereas the region between b_m and c_m belongs 'purely' to the Fuzzy Set.

The method proposed in this thesis consists of the determination of the parameters a_m, b_m, c_m and d_m, for the different Fuzzy Sets of a variable, based on the changes of convexity in the corresponding HSC. The values which define each of this parameters were placed at the minimum and the maximum values of the range covered by the curve, and in each point of the HSC presenting a change in the curve character, i.e., from concave to convex or vice versa. All of those values are hereafter called breaks of the curve.

The MF of the Fuzzy Sets of each variable were constructed by firstly defining b_m and c_m. Thus, b_m and c_m of the central Fuzzy Set were defined by the values of the breaks in the curve comprising the maximum suitability for the involved variable. This is the part of the curve where the slope equals or approaches zero and hereafter referred to as the plateau. These breaks also defined the d_{m-1} and the a_{m+1} parameters of the adjacent Fuzzy Sets located at the left and right of this Fuzzy Set, respectively. The a_m and the d_m of the central Fuzzy Set were defined by the following breaks both right and left of the plateau and also correspond to the parameters c_{m-1} and b_{m+1} of the adjacent Fuzzy Sets located at the left and right. The further breaks in both sides, the left and the right, corresponding then to the b_{m-1} and c_{m+1}, etc. The procedure to define the Fuzzy Sets parameters continued until a MF parameter was set for each break of that curve (Fig. 6 - A). Finally, the outer values of the outer MF were set equal to the last value to obtain a trapezoidal MF. This procedure was applied for the remaining HSCs constituting the first part of the fuzzy inference system.

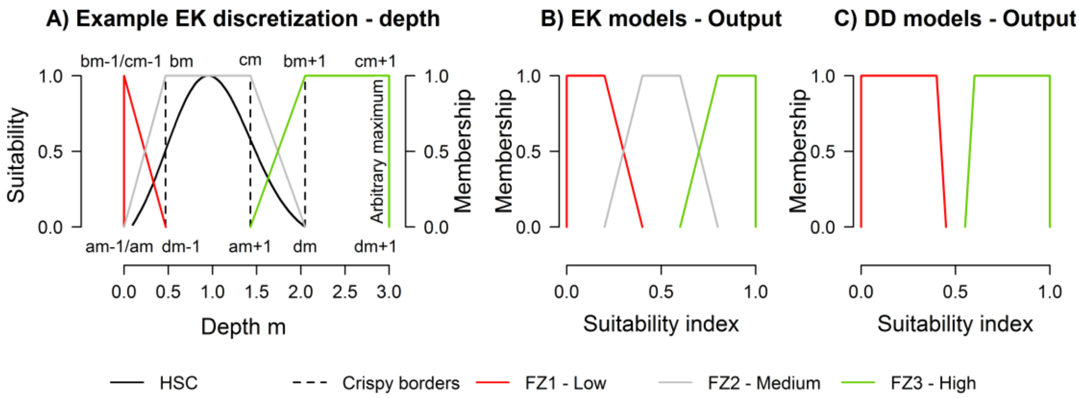

Fig. 6 Example of discretization of the variable depth in three Fuzzy Sets for the adult brown trout. The central Fuzzy Set covers the maximum suitability. The following interval in both sides is fuzzy and shared with the adjacent Fuzzy Sets. The process continues alternating defined regions with fuzzy regions until no more breaks are available for the analyzed HSC. EK = Expert-knowledge, DD = Data-driven.

Instead of the procedure to determine the number, shape and parameters of the Fuzzy Sets of the input variables, the output variable was theoretically determined. Te main objective was to produce a model with an output between 0 and 1, like the Habitat Suitability Index (HSI); zero means unsuitable whereas one means the maximum suitability. The output variable was discretized in three Fuzzy Sets: Low, Medium and

High (Fig. 6 - B). The output variable was divided in five uniform intervals of range 0.2, resulting in alternating intervals that either fully belong to a Fuzzy Set or that cover the transition between two adjacent Fuzzy Sets.

Once the Fuzzy Sets were created, Fuzzy Rules were defined based on the information derived from the HSC following the next procedure. The first step consisted of the assignment of a partial suitability to each of the considered Fuzzy Sets, independently for each variable. The partial suitability for the Fuzzy Set that covered the plateau of the HSC corresponded to High suitability, whereas the partial suitability of the Fuzzy Sets on the extremes was Low. The partial suitability assigned to the remaining Fuzzy Sets was Medium thus the suitability of each HSC provided the basic frame to assign the corresponding partial suitability. Then the partial suitability values of the three variables were aggregated in a single suitability value becoming the rule consequent. The combined suitability output for each rule was determined following the following criteria. If the depth was extremely Low or extremely High the output of the rule including that Fuzzy Set was always Low and could not be compensated by any better suitability output from the remaining variables. If the velocity was extremely high the combined suitability output was always Low and could not be compensated by any better partial suitability output from the other two variables. The combined suitability output for the remaining rules was determined as the maximum appearance, independently of any better remaining partial suitability output. For instance, if velocity suitability was High, depth suitability was Medium and substrate suitability was Medium the combined suitability was Medium, neglecting the high suitability from the velocity variable. If a draw appeared the suitability for that rule was determined as Medium in any case, if there was no conflict with the initial assumptions. Several authors considered depth a non-controlling variable regarding the older brown trout size classes (Bovee, 1978; Ayllón *et al.*, 2010) thus the aforementioned constrain was relaxed only for large brown trout, and the largest depth was assigned to Medium suitability if the combination of partial suitability was not in conflict with the remaining assumptions.

3.5 Data-driven fuzzy approach for habitat suitability modelling

3.5.1 Sub-sampling

Prevalence can have a strong effect on model performance (Manel, 2001). The prevalence for every river database was extremely low in our study, because the number of data collected to characterise habitat availability was several times the number of habitat use the highest prevalence among study sites was 0.15. To avoid undesirable effects, a sub-sampling procedure was applied; the main objective was to obtain a new database with 0.5 prevalence but statistically similar to their originals. It should be done with a multivariate procedure to keep the combinations of velocity, depth and substrate.

For every river database the sub-sampling methodology followed the next procedure. First the Euclidean distance of each case to the centre of gravity of the Availability dataset was calculated, with the centre of gravity calculated as: Centre of gravity (CDG) = (Average Velocity, Average Depth, Average Substrate Index). An example for the Senia-2006 dataset is shown in the Fig. 7. These distances comprise in a unique and simple index the three microhabitat variables. The records with little distances in general have more common values of the three variables (i.e. more common values of Velocity, Depth and Substrate index) than the records with larger distances. Then, a cumulated frequency histogram of these distances was generated and then the records were extracted in a systematic sampling procedure through that cumulated frequency distribution. The number of sub-sampled records in each Availability dataset was the number of records in its respective Use dataset, so the resulting prevalence was 0.5.

Generally, the histogram showed a steep slope for low distance values (i.e. many values are close to the Centre of gravity), and the slope decreased asymptotically as the distance increased. Regarding the sub-sampled dataset, the value triads (Velocity, Depth, Substrate) of the initial extracted records, in general, are more common than the last extracted triads that are rarer. Therefore, the subsample has each variable distribution similar to the original Availability dataset. The Fig. 7 shows an example of the selection of the records suitable to be extracted for the Senia-2006 dataset where 13 records were sub-sampled. It is important to remember at this point, the survey was done with an equal effort surveying methodology; otherwise a strong bias would be committed.

Statistical tests were then applied to check differences between the original Availability dataset and its respective sub-sample. These tests were applied to the three microhabitat variables separately. The applied tests were a robust generalization of Welch test (Welch, 1951), which compares means, and a robust generalization of Kruskal-Wallis test (Rust and Filgner, 1984), which is a non-parametric test on variance analysis.

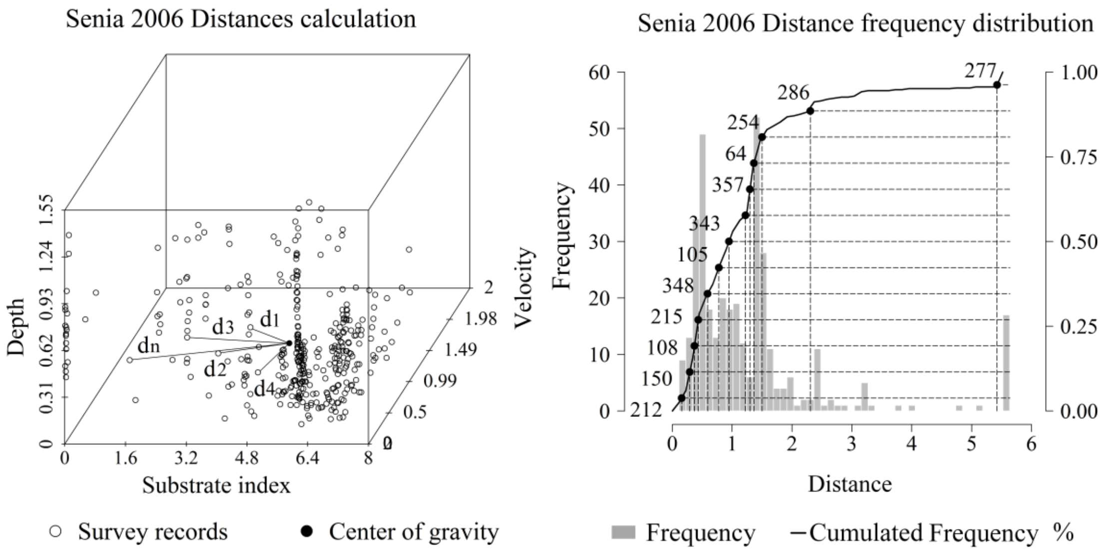

Fig. 7 Steps followed during the sub-sampling approach in order to generate 0.5 prevalence databases (source: Muñoz-Mas *et al.*, 2012). On the left, calculation of the distance of every case to the centre of gravity. On the right cumulative frequency analysis and systematic extraction of the corresponding cases. Numbers next to the dots correspond to the extracted cases.

3.5.2 Data-driven fuzzy approach for habitat suitability modelling

The development of the Data-driven fuzzy models followed the methodology presented by Mouton (2008). Their development involve two main procedures similarly to the steps followed to develop the Expert-knowledge fuzzy models explained above. Firstly the optimization of the Fuzzy Sets is carried out and then the Fuzzy Rules are optimized.

The optimization of the Fuzzy Sets aims at the optimal discretization of the input variables in categories; Low, Medium, High etc. based on the Shannon–Weaver entropy (Shannon and Weaver, 1963). The main goal was to obtain a balanced discretization based on the number of cases included in each Fuzzy Set, in order to improve the results in the optimization of Fuzzy Rules; otherwise, if a Fuzzy Set of an input variable contains very few training instances, the rules that apply this Fuzzy Set will be poorly trained. In the Data-driven approach, the Fuzzy Set geometry is also defined by its membership function and, following the same criteria than the Expert-knowledge approach, the selected geometry was the trapezoidal. Likewise the EK methodology, four parameters, a_m, b_m, c_m and d_m, determine the degree of membership of a given value to that Fuzzy Set. The optimization of the Fuzzy Sets consists of the slight modification of these parameters, step by step; after each modification the Shannon–Weaver entropy following (1) is calculated:

$$entropy = -\frac{1}{log_2 n} \sum_{i=n}^{n} p_i \, log_2 p_i \quad (1)$$

where n is the number of classes or Fuzzy Sets and p_i the proportion of data belonging to the class i. These steps are going on while an improvement of the entropy is obtained and until the entropy reach the threshold or the maximum. In order to calculate the proportion of data in a given class, a datum is assigned to a given class (Fuzzy Set) if its membership is higher than 0.5 (see Mouton *et al.* 2008 for further details).

A Presence/Absence was the selected output discretization thus two Fuzzy Sets were generated to cover the output range. Although more gradual discretizations are sometimes used, the method of direct observation for individual fish lead to this option, preventing from any other discretization of the output, because the survey was carried out trout by trout and there was no micro-scale information about fish abundance (which usually demands an estimation at larger scales).

Therefore, the habitat suitability was discretized in two Fuzzy Sets with no overlapping areas. The absence of overlapping areas do not provide always an integer value, zero for Absence or one for Presence, because the final suitability depends on the centre of gravity of both Fuzzy Sets, considering the areas under their respective degrees of fulfilment. Therefore, an smooth transition from presence to absence is obtained.

The Data-driven fuzzy models also present a set of rules which relate the inputs with the output variable. These rules are also constructed as an If-Then sequence, and the optimization of the proper consequent for every set of antecedents was the main goal of the following step. This optimization was carried out based on the information contained in the pooled database for each size class (i.e. the pooled data from the Use dataset and the sub-sampled Availability).

The optimization was done with the software FISH (Mouton 2010). During the optimization process FISH© executes a defuzzyfication procedure generating a fuzzy classification (Mouton *et al.*, 2008) and the entire optimization is based on the comparison of the potential output and the measured one by means of a performance criterion based on the confusion matrix (Mouton *et al.*, 2010a). The entire optimization was based on the performance criterion of Cohen's Kappa (Cohen, 1960) (hereafter Kappa) which showed acceptable results in previous studies (Mouton *et al.*, 2008; Muñoz-Mas *et al.*, 2012). The Fuzzy Rules were optimized based on the hill-climbing algorithm (Michalewicz and Fogel, 2000) in FISH. For each fuzzy rule (or set of antecedents) the process start at one random consequent (for example Low), then this consequent is changed to its adjacent category (for example, Medium) and the Kappa is calculated. If the model performance increases in the current step, the algorithm continues with the adjusted rule; if not, it retains the previous one repeating the process with a new output category.

To assess the model convergence and robustness, for the adult and the juvenile classes, 10 times three-fold cross validations were done, with five iterations each; for fry, 10 times two-fold cross validations were done, also with five iterations each. The value of Kappa was then calculated as the average value of such criterion in the 150 or the 100 resulting confusion matrices. The optimal consequent of a rule was the consequent that occurred with the highest frequency in the optimizations. The rules that did not present any case to be trained (hereafter 'uncovered rules') were assessed by comparison with the Expert-knowledge approach.

3.6 Validation, model adjustment and models comparison

A spatially explicit validation was carried out in order to test models generalization and their transferability over the study site of the Cabriel River. The validation was carried out in terms of model sensitivity and model specificity (i.e. the ability of the generated models to assess fish location with the maximum suitability, but not assessing the entire reach with the maximum suitability). The flow present in the validation survey (0.89 m^3/s) was simulated in the hydraulic model and the habitat suitability for each pixel was calculated using the generated models, i.e. Expert-knowledge and Data-driven models. The entire assessment was carried out with the fish habitat module of the CASiMiR numerical modelling toolbox (Jorde, 2000; Schneider, 2001) discretising the modelled area in pixels of 1 m^2.

A frequency analysis, pixel by pixel, of the assessment of the entire area for the nine generated models was carried out, after reclassifying suitability in five equal intervals (width 0.2). In addition, the frequency analyses of suitability at the fish locations (for each size class) were carried out applying the same discretization; then both distributions were compared.

It is relevant to remember that the Expert-knowledge fuzzy models were based on the HSCs which were developed applying an smoothing technique over the calculated histograms. This procedure filled the intervals where there were no fish observations, in order to provide unimodal curves. This filling-up modified the curve values toward higher values of suitability; however, the Data-driven approach does not allow automatically the implementation of similar modification. Therefore, the irregular distribution of the Use data can produce irregular patterns on the transitions between suitable and unsuitable conditions gathered in the Fuzzy Rules set.

To reduce or mitigate this effect, the output of the assessment of the trout locations by means of the generated Data-driven models was used as a feedback to modify the corresponding Fuzzy Rules in the Data-driven approach. Accordingly to the procedure applied in the Expert-knowledge approach, the modifications tried to maximize the percentage of fish locations in pixels assessed with high or the maximum suitability, thus surely producing overpredictive model. However overprediction should not be considered a model error (Mouton *et al.*, 2010b). The absence of the target species in a suitable area may be due to the unbalanced colonization of habitats in the study area given the presence of barriers, temporal population variations or sampling inefficiencies (MacKenzie *et al.*, 2003). Additionally, the modification was intended to be the minimum in order to keep the results in the data driven approach, with minimum human intervention.

The number of Fuzzy Sets and Fuzzy Rules can vary regarding the considered approach. Additionally the fact that the Use data were collected for individual fish limited the Data-driven output, reducing the problem to a Presence/Absence approach thus the considered output in the Expert-knowledge approach and in the Data-driven approach differed in the number of fuzzy sets. Altogether hampered the models comparison, thus two different approaches were carried out. The first one based on similarities and differences on the rules consequents and the second one based on the models performance. The first comparison could be considered more mathematical, and the second more biological.

It has been demonstrated that Fuzzy Rules and their corresponding Membership functions have a strong impact on model training (Mouton, 2008) and hence on model performance thus the direct comparison of the Fuzzy Rules derived from different Fuzzy Sets and approaches; Expert-knowledge and Data-driven, was considered potentially imprecise. Then in order to check the robustness of the developed Expert-knowledge Fuzzy Rules, the Fuzzy Rules based on Expert-knowledge Fuzzy Sets were optimized by means of the Data-driven approach. The Fuzzy Sets generated by means of the Data-driven approach and the Expert-knowledge approach did not match thus its counterpart was unfeasible. Then, the consequents were compared and discussed between Data-driven and Expert-knowledge models for each fish class.

The second comparison was carried out between the models performance with the simulated flow at the validation site. The assessment of the four developed models for each size class were pairwise compared in the spatially explicitly context of the river reach. However, the outputs from the two main approaches dissuaded from a pixel by pixel comparison, thus the spatial explicit comparison was carried out with the Map

Comparison Kit version 3.2.2 (Research Institute for Knowledge Systems, The Netherlands, 2011). This software allows the comparison taking into account certain degree of tolerance between the categories of the overlaid pixels and taking into account the surrounding pixels. The comparison becomes a fuzzy comparison in both the category definition and in the considered location (Hagen, 2003; Hagen-Zanker *et al.*, 2005).

The fuzziness of the overlaid categories was implemented by assigning to each cell a membership vector instead of a single category. Each element in the vector declares the degree of membership for one category (Hagen, 2003) and ranges from one, perfect agreement, to zero, null agreement. All this information is gathered in the 'category similarity matrix', where similarity between categories decreases when distance from the diagonal increases. The considered decrease was linear from the same category, with perfect agreement, to the further category, with null agreement, presenting perfect consistency (Saaty, 1980) (Table 3).

Table 3. Category similarity matrix used in the spatial fuzzy comparison of the models. The considered similarity linearly decreases as the category becomes farther. The linearity provided a perfect consistency (Saaty, 1980).

		Suitability				
		0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
Suitability	0.0-0.2	1	0.75	0.5	0.25	0
	0.2-0.4	0.75	1	0.75	0.5	0.25
	0.4-0.6	0.5	0.75	1	0.75	0.5
	0.6-0.8	0.25	0.5	0.75	1	0.75
	0.8-1.0	0	0.25	0.5	0.75	1

In addition, the fuzziness of location considers the area surrounding each pixel, therefore the fuzzy representation of a given pixel depends on the cell itself and, to a lesser extent, on the cells within a certain distance in its neighbourhood (Hagen, 2003). The extent to which the neighbouring cells influence the fuzzy calculation is represented by a distance decay function. There is no a standard universal way to select the proper decay function and its parameters (personal communication Hagen, 2012). The brown trout could be basically considered a territorial fish (Chapman and Bjonn, 1969; Titus, 1990; Johnsson *et al.*, 2000) thus its distribution along the validation site was expected to be determined by its territoriality (i.e. the disposition of the surrounding individuals conditioned the position of the considered individual). Therefore the selected extent of the influencing neighbour cells was based on the distances of each individual to the nearest. Therefore, in the present study, a linear decay (cone shape, defined by slope = 1) and a variable radius of 5 m, 5 m and 2.5 m for adult, juvenile and fry, respectively, were considered as appropriate accordingly to the mean distance to the nearest trout of the corresponding size class. The fuzzy Kappa statistic was selected as the similarity index. The fuzzy Kappa statistic is similar to the traditional Cohen's Kappa (Cohen, 1960), correcting the overall agreement of both models by the agreement expected to occur by chance.

The results of the previous analyses provided us with a general assessment on the performance of the models over the validation area, but they did not consider trout densities. Trout could be considered a territorial fish (Chapman and Bjonn, 1969; Titus, 1990; Johnsson *et al.*, 2000), thus a validation considering fish density was appropriate because the correct assessment of the most populated areas could be considered a keystone in the selection of a given model for further analysis. An estimated trout density would be desirable to carry out the proper validation, but the survey methods do not allow its correct calculation. Therefore, the trout density was calculated as the number of fish observed of a given class per unit area. Fish density was calculated using the tool Kernel Density in ESRI® ArcMapTM 9.3 (Copyright© 1999-2008 Esri Inc.) with a radius equal to the mean distance to the nearest trout of the corresponding class accordingly to the aforementioned calculation of the fuzzy comparison.

This density value was standardized between 0 and 1 and discretized, also in five intervals corresponding to; Very Low density, Low density, Medium density, High density and Very high density. The mean habitat suitability, assessed by means of the generated models, was calculated and the results were compared and plotted (Fig. 9). Finally, the variability of the assessment for the areas with similar density was also analysed by plotting the range of the assessed suitability in terms of maxima and minima.

4 Results

4.1 Habitat Suitability Curves and Fuzzy Sets - Adult

4.1.1 Habitat Suitability Curves - Adult

The Category II ½ curves for the adult trout showed the maxima at 0.125 m/s, 0.35 m and 6 (corresponding to cobble) for velocity, depth and substrate index respectively (Fig. 8). The Category III curves showed a clear displacement of the highest suitability to higher velocity and larger depth, thus the maximum suitability appeared around 0.8 m/s and 1.3 m respectively, and 5.5 for substrate index (corresponding to gravel-cobble). Only the depth curve showed a clear pointed shape, but the other Category III curves were wider than the Category II ½ counterparts (Fig. 8).

4.1.2 Expert-knowledge Fuzzy Sets - Adult

With the Expert-knowledge approach, based on Category II ½ and Category III curves, three Fuzzy Sets were produced for the variable velocity, four Fuzzy Sets for the variable depth and two Fuzzy Sets for substrate, presenting certain similarity in the shapes of the Fuzzy Sets but varying in their partial suitability (Fig. 8).

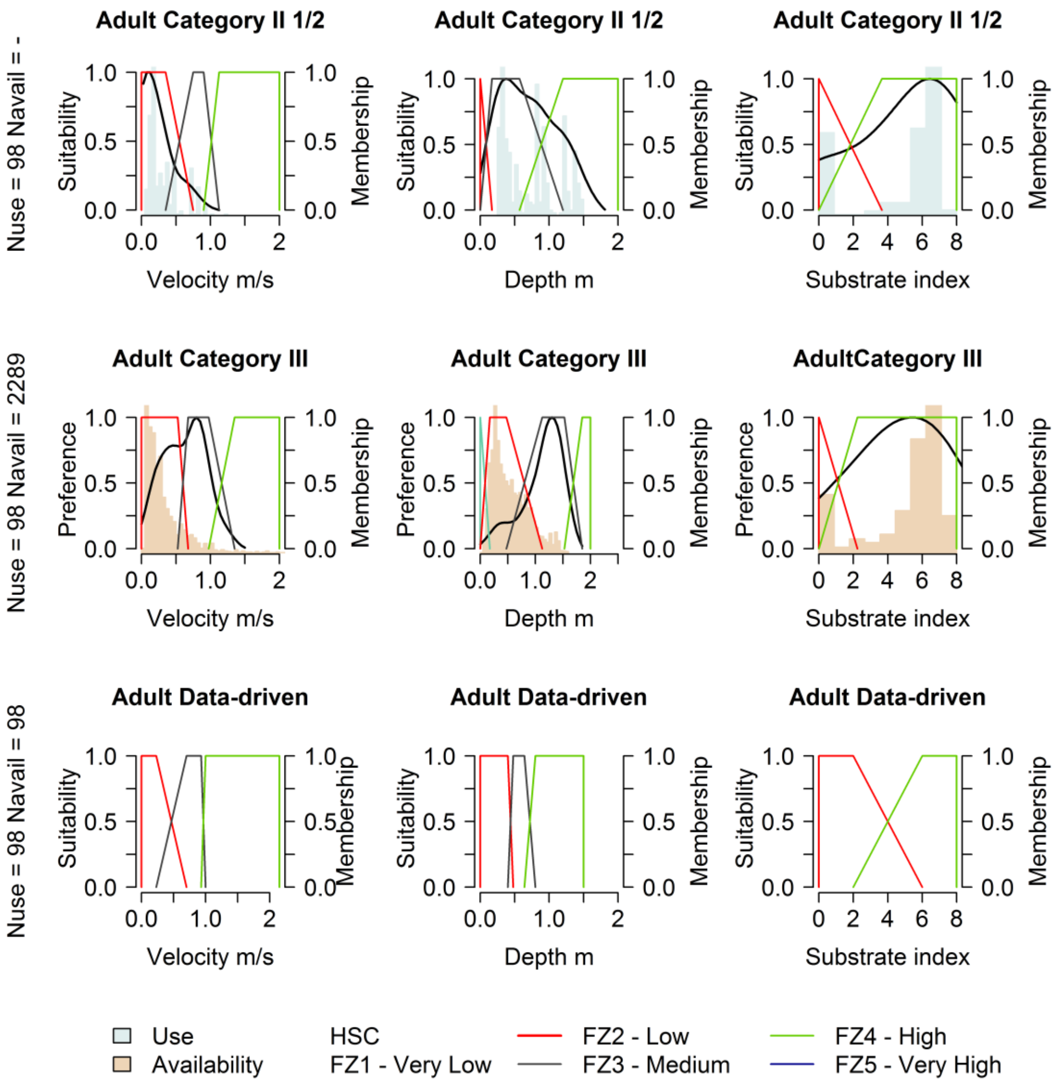

Fig. 8. Category II ½ and Category III Habitat Suitability Curves and their corresponding Fuzzy Sets for Adult brown trout. The last sequence corresponds to the Data-driven Fuzzy Sets obtained from the Shannon-Waver entropy based optimization. Nuse correspond to the amount of trout observations whereas the Navail to the amount of observation about the surrounding conditions considered in the development of the corresponding model. In the back the complete frequency analysis of the Use data and the Availability data is shown.

4.1.3 Data-driven Fuzzy Sets - Adult

After the sub-sampling, the statistical tests did not show significative differences between the original database and the extracted sub-sample (Table 4). The Data-driven approach discretised the variables velocity and depth in three Fuzzy Sets, and the variable substrate in two Fuzzy Sets, achieving the Shannon–Weaver entropy values of 0.42, 0.99 and 0.94 for velocity, depth and substrate respectively (Fig. 8).

Table 4 Test results for every sub-sampled Availability dataset. Tval. means test value and S lev. means signification level.

River	Year	Robust generalization of Welch test						Robust generalization of Kruskal-Wallis					
		Depth (m)		Velocity (m/s)		Substrate (S)		Depth (m)		Velocity (m/s)		Substrate (S)	
		T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.
Cuervo	2009	0.14	0.71	0.4	0.52	0.66	0.42	0.08	0.76	0.36	0.54	0.58	0.44
Guadiela	2009	0.02	0.88	0.79	0.38	0.03	0.86	0	0.98	0.25	0.61	0.04	0.85
Senia	2007	0.02	0.87	0.78	0.38	-	-	0	0.98	0.25	0.61	-	-
Jucar	2007	0.05	0.82	0.26	0.52	0.47	0.51	0.04	0.84	0.06	0.8	0.85	0.35

4.1 Habitat Suitability Curves and Fuzzy Sets - Juvenile

4.1.1 Habitat Suitability Curves - Juvenile

The juvenile case produced Category II ½ curves with maxima at 0.175 m/s, 0.35 m and substrate index around 6 (corresponding to cobble) for the variables velocity, depth and substrate respectively (Fig. 9). The Category III curves showed a certain displacement of the highest suitability to higher velocity and larger depth, achieving the maxima at 0.55 m/s and 0.55 m respectively, and the highest suitability for substrate index of 4 (i.e. gravel). In comparison with the adult, the Category III curves for juvenile did not present as dramatic displacements; regardless the considered variable, these curves presented a wider suitability than the Category II ½ counterparts (Fig. 9).

4.1.2 Expert-knowledge Fuzzy Sets - Juvenile

The Expert-knowledge approach showed a different discretization in the models derived from Category II ½ and Category III. From Category II ½ curves, two, four and two Fuzzy Sets were obtained for velocity, depth and substrate, respectively, whereas from the Category III, the Fuzzy Sets presented similar discretization as the adult, with three, three and two Fuzzy Sets for velocity, depth and substrate (Fig. 9).

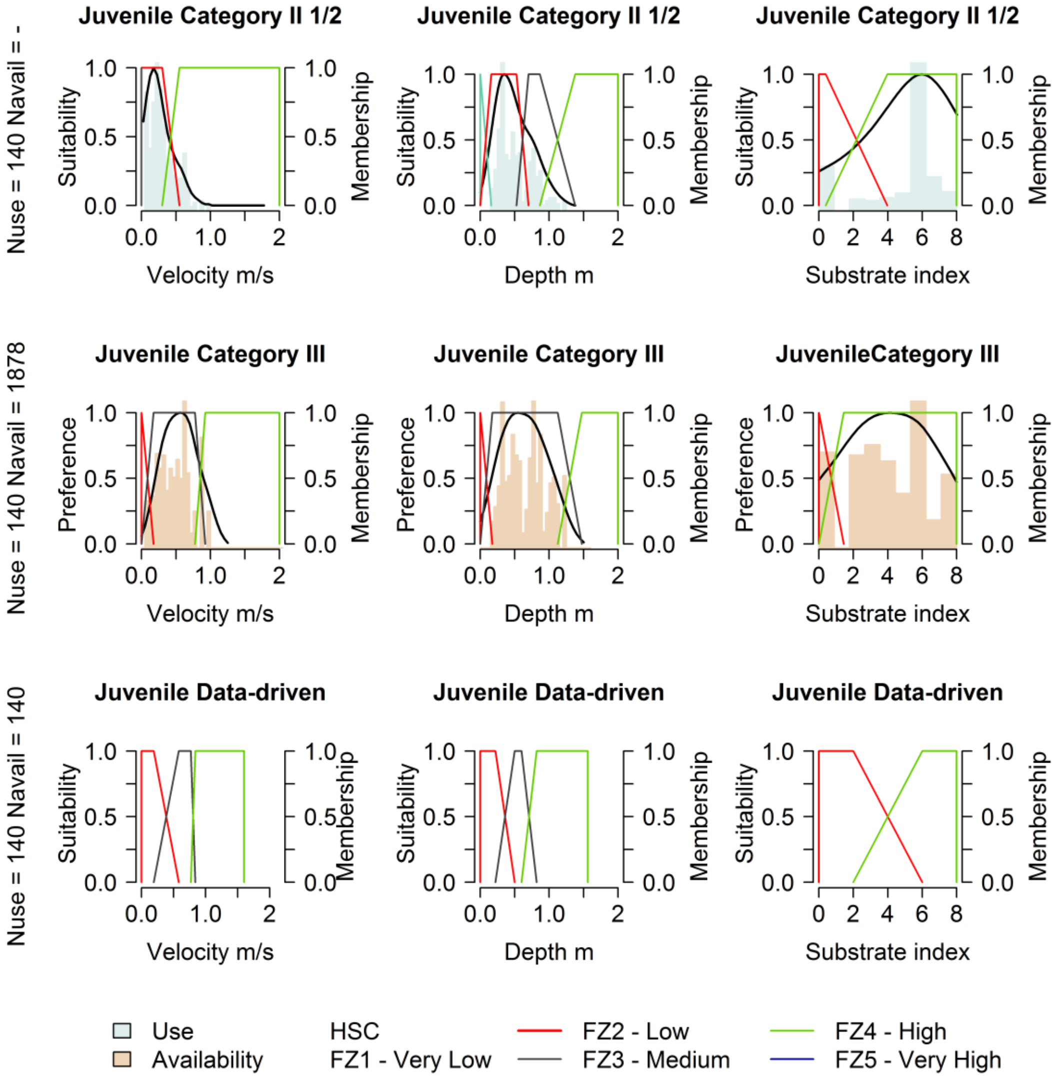

Fig. 9 Category II ½ and Category III Habitat Suitability Curves and their corresponding Fuzzy Sets for Juvenile brown trout. The last sequence corresponds to the Data-driven Fuzzy Sets obtained from the Shannon-Waver entropy based optimization. Nuse correspond to the amount of trout observations whereas the Navail to the amount of observation about the surrounding conditions considered in the development of the corresponding model. In the back the complete frequency analysis of the Use data and the Availability data is shown.

4.1.3 Data-driven Fuzzy Sets - Juvenile

As in the case of large fish, the sub-sampling methodology did not produced statistical differences between the original database and the subsample, and the resulting database was considered suitable for further analysis (Table 5).

Table 5. Test results for every sub-sampled Availability dataset. T. val. means test value and S. lev. means signification level.

River	Year	Robust generalization of Welch test						Robust generalization of Kruskal-Wallis					
		Depth (m)		Velocity (m/s)		Substrate (S)		Depth (m)		Velocity (m/s)		Substrate (S)	
		T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.
Senia	2006	0.07	0.79	1.24	0.28	0.02	0.9	0.25	0.62	0.05	0.83	0.02	0.9
Senia	2007	0.03	0.87	2.09	0.17	0.19	0.67	0.08	0.78	2.34	0.13	0.01	0.91
Jucar	2006	0.06	0.81	0.23	0.64	0.98	0.34	0.05	0.83	1.25	0.26	1.06	0.3
Jucar	2007	0.1	0.75	0.03	0.87	0.43	0.52	0.15	0.7	0.25	0.62	0.62	0.43
Cabriel	2005 Set	0.04	0.83	0.31	0.58	0.04	0.84	0.02	0.89	0.07	0.79	0.01	0.94
Cabriel	2005 Oct	0.47	0.5	0.17	0.68	0.52	0.47	0.3	0.58	0.06	0.81	0.56	0.45

The Data driven approach discretized the variable velocity in three Fuzzy Sets, the depth in three, and substrate in two, achieving the Shannon–Weaver entropy values of 0.62, 0.98 and 0.87 respectively (Fig. 9).

4.2 Habitat Suitability Curves and Fuzzy Sets - Fry

4.2.1 Habitat Suitability Curves - Fry

The fry showed Category II ½ unimodal curves with maxima at 0.12 m/s, 0.25 m and substrate index of 0 (corresponding to silt and vegetation) for the variables velocity, depth and substrate respectively (Fig. 10). The Category III curves presented certain displacement toward larger values of velocity and depth, thus the peak of the curve appeared at 0.55 m/s and 0.57 m respectively, whereas the highest suitability for substrate remained constant (Fig. 10).

4.2.2 Expert-knowledge Fuzzy Sets - Fry

The Expert-knowledge approach lead to a different number of Fuzzy Sets for the models derived from the Category II ½ and the Category III curves. The Category II ½ curves produced fewer Fuzzy Sets than the Category III. The variable velocity was discretized in two Fuzzy Sets and the variable depth in three Fuzzy Sets whereas the variable substrate presented two Fuzzy Sets. The discretization derived for the Category III curves was similar for the three involved variables presenting three Fuzzy Sets for the three variables (Fig. 10).

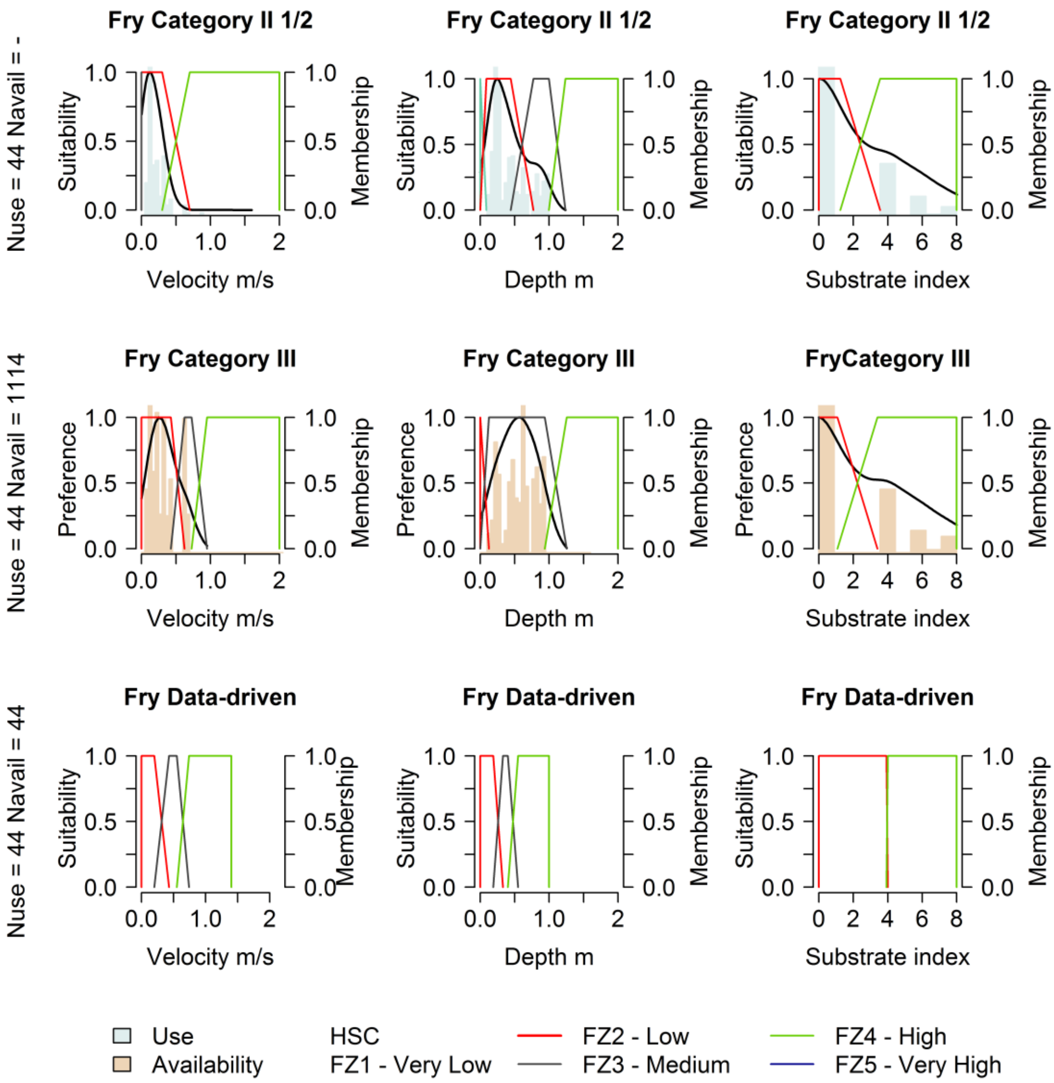

Fig. 10 Category II ½ and Category III Habitat Suitability Curves and their corresponding Fuzzy Sets for Brown trout fry. The last sequence corresponds to the Data-driven Fuzzy Sets obtained from the Shannon-Waver entropy based optimization. Nuse correspond to the amount of trout observations whereas the Navail to the amount of observation about the surrounding conditions considered in the development of the corresponding model In the back the complete frequency analysis of the Use data and the Availability data is shown.

4.2.3 Data-driven Fuzzy Sets - Fry

The sub-samplig methodology did not produced statistical differences between data sets (Table 6). The Data driven approach discretized the variable velocity in three Fuzzy Sets, depth in three Fuzzy Sets and substrate in two Fuzzy Sets, achieving the Shannon–Weaver entropy values of 0.59, 0.98 and 0.98 for the variables velocity, depth and substrate respectively (Fig. 10). Notice the variable substrate presented almost no overlapping.

Table 6. Test results for every sub-sampled Availability dataset. T. val. means test value and S. lev. means signification level.

River	Year	Robust generalization of Welch test						Robust generalization of Kruskal-Wallis					
		Depth (m)		Velocity (m/s)		Substrate (S)		Depth (m)		Velocity (m/s)		Substrate (S)	
		T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.	T.val.	S.lev.
Turia	2006	1.17	0.29	0.82	0.37	0.87	0.37	2.2	0.1	0.7	0.4	0.7	0.4
Jucar	2007	1.17	0.29	0.82	0.37	0.83	0.37	2.2	0.1	0.7	0.4	0.7	0.4

4.3 Fuzzy Rules

4.3.1 Expert-knowledge Fuzzy Rules

The Expert-knowledge Fuzzy Rules were generated accordingly to the aforementioned methodology and are summarized in the Table 7. The Expert-knowledge approach allowed the definition of the whole Fuzzy Rules set thus no uncovered rules appeared in the application of this approach.

4.3.2 Data-driven Fuzzy Rules

The Data-driven optimization of the Fuzzy Rules achieved the values 0.31 ± 0.04, 0.21 ± 0.02 and 0.37 ± 0.08 of the Cohen's Kappa for the adult, the juvenile and the fry size class respectively. The training for the adult model presented five rules uncovered (with no cases to be trained) and the Expert-knowledge approach was used in their determination. The juvenile model presented five uncovered rules and were also determined using expert knowledge. The fry model had more uncovered rules, (seven), and similarly the Expert-knowledge approach was used in their definition (Table 7).

Table 7 Summary of the Fuzzy Rules for the habitat suitability models. EK = Expert-knowledge, DD = Data-driven. Italics mean adjusted/modified rule. Asterisk (*) means uncovered rule. The output was determined through authors consensus in the EK models. Bold means discrepant output in the two approaches. Numbers refer to the amount of cases for training each rule.

Rule Nº	Velocity m/s	Depth m	Substrate s	EK Adu Cat II 1/2	DD Adu Cat II 1/2	EK Adu Cat III	DD Adu Cat III	DD Adult	EK Juv Cat II 1/2	DD Juv Cat II 1/2	EK Juv Cat III	DD Juv Cat III	DD Juvenile	EK Fry Cat II 1/2	DD Fry Cat II 1/2	EK Fry Cat III	DD Fry Cat III	DD Fry
1	L	VL	L			L	*		L	L1				L	*			
2	L	L	L	L	*	M	L22	L	**H**	**L15**	L	L1	L	H	H24	L	L1	L
3	L	M	L	**H**	**L30**	M	L3	L	M	L30	M	L13	L	H	H10	H	H28	H
4	L	H	L	M	L1	M	*	H	L	L3	L	*	*L/H*	L	*	L	*	H
5	M	VL	L			L	*		L	*								
6	M	L	L	L	*	M	L2	H	M	H2	L	*	L			L	*	L
7	M	M	L	M	L3	H	*	H	M	*	H	H27	H			H	H6	L
8	M	H	L	M	*	M	*	H*	L	*	L	*	H*			L	*	H*
9	H	VL	L			L	*		L	*				L	*			
10	H	L	L	L	*	L	*	H	L	*	L	*	L	L	L1	L	*	L*
11	H	M	L	L	*	L	*	H*	L	*	L	*	L*	L	*	L	*	L*
12	H	H	L	L	*	L	*	H*	L	*	L	*	L*	L	*	L	*	L*
13	L	VL	H			L	*		L	L1				L	*			
14	L	L	H	L	*	M	L109	*L/H*	H	H104	L	L1	H	**H**	**L37**	L	L2	L
15	L	M	H	H	H115	H	H49	H	H	H51	**H**	**L44**	H	M	H7	**H**	**L21**	*L/H*
16	L	H	H	M	H37	M	*	H	L	L9	L	L2	H	L	*	L	*	H
17	M	VL	H			L	*		L	*								
18	M	L	H	L	*	**H**	**L9**	L	H	H39	L	*	H			L	L1	H
19	M	M	H	H	H7	H	H1	H	M	*	H	H172	H			M	L26	L
20	M	H	H	M	*	M	*	H*	L	*	L	L1	H*			L	*	H*
21	H	VL	H			L	*		L	*				L	*			
22	H	L	H	L	*	L	L1	L	L	L15	L	*	L	**L**	**H7**	L	*	L*
23	H	M	H	L	L2	L	*	H*	L	*	L	L9	L	L	*	**L**	**H1**	H
24	H	H	H	**L**	**H1**	L	*	H	L	*	L	*	L*	L	*	L	*	L*

4.4 Fuzzy Rules robustness and Fuzzy Rules comparison

The training of the Fuzzy Rules using the Data-driven approach and considering the Fuzzy Sets based on the HSCs produced a model for adults with a Kappa value of 0.20

± 0.03 after the Category II ½ curves, and 0.21 ± 0.03 after the Category III curves. The juvenile models obtained Kappa = 0.18 ± 0.03 (based on Category II ½) and Kappa = 0.19 ± 0.05 (based on Category III). The fry models presented the highest values of Kappa, being 0.39 ± 0.1 and 0.37 ± 0.1 for the Category II ½ and the Category III-based models, respectively.

Therefore, the performance indices were in any case lower than those obtained with the Fuzzy Sets developed within the Data-driven approach. The amount of uncovered rules was larger than in the pure Data-driven approach. The adult models presented 55 % and 66 %, based on Category II ½ and Category III curves, respectively. The juvenile model presented and slightly improvement with 54 % and 50 % of uncovered rules, whereas the fry had a higher amount of uncovered rules, i.e. 62 % and 55 % in the two models. Accordingly, these results demonstrated the improvement of the proper distribution of the training cases on the optimization results (i.e. higher values of Kappa) and the reduction of uncovered rules.

The Expert-knowledge approach discretized the output in three Fuzzy Sets (i.e Low suitability, Medium suitability and High suitability) and the Data-driven approach in two Fuzzy Sets (i.e Low suitability and High suitability), thus the comparison of the output meant a significant difference only if an approach provided High suitability while the other provided Low suitability.

4.4.1 Fuzzy Rules comparison - Adult

The Data-driven Fuzzy Rules for the adult, based on the Category II ½ Fuzzy Sets, differed in two rules with its Expert-knowledge counterpart (Rule #3 and Rule #24 Table 7). However the major difference appeared in a single rule (Low Velocity, Medium Depth and Low Substrate) because this presented a large amount of training cases (30 cases) whereas the other discrepant rule appeared in the rule covering High Velocity, High Depth and High Substrate which was trained based on a single observed trout (1 case) thus its discrepancy was considered uncertain in comparison with the previous case.

The Expert-knowledge approach based on Category III curves discretized the variable depth in four fuzzy sets (i.e. Very Low depth, Low depth, Medium depth and High depth). The Data-driven fuzzy model trained with that Fuzzy Sets based on the Category III HSCs did not presented any case to train the rules including the Very Low depth Fuzzy Set. Additionally it presented a single discrepant rule corresponding to Medium Velocity, Low Depth and High Substrate; its suitability was considered High in the Expert-knowledge approach and Low in the Data-driven approach but it was trained with only 9 cases (Rule #18 Table 7). However an appreciable difference was observed in the assessment of the rules that included the Fuzzy Set corresponding to Low Depth, thus the Expert-knowledge approach showed permissive assigning the Medium suitability to that rules whereas the Data-driven approach assessed that rules always as Low thus restricting the suitable areas to deeper areas in comparison with the Expert-knowledge approach.

4.4.2 Fuzzy Rules comparison - Juvenile

The comparison of the Fuzzy Rules between Expert-knowledge and Data-driven based on HSCs, for juvenile, showed discrepancies in only one rule between models based on Category II ½ HSCs (Rule #2 Table 7) and similarly one rule differed in the model based on Category III HSCs (Rule #15 Table 7). The Category II ½ based models differed in the rule; Low Velocity, Low Depth and Low Substrate, which was considered

High in the Expert-knowledge approach and Low in the Data-driven. In the Category III based models differed in the rule; Low Velocity, Medium Depth and High Substrate, thus it was considered High in the Expert-knowledge approach and Low in the Data-driven (Table 7).

4.4.3 Fuzzy Rules comparison - Fry

The fry presented the largest differences, with two discrepant outputs in the Category II ½ case (Rule #14 and Rule #22 Table 7) and two in the Category III case (Rule #15 and Rule #23 Table 7). However the major difference in the Category II ½ case appeared in a single rule, Low Velocity, Low Depth and High Substrate, trained with 37 cases which corresponded to the 44 % of the whole database; based on Category III curves, the major difference was in the rule; Low Velocity, Medium Depth and High Substrate, trained with 21 cases (25 % of the whole database).

4.5 Fuzzy models performance

4.5.1 Fuzzy models performance - Adult

The assessment of the simulated flow at the validation site showed different results for the models of adult brown trout. The Expert-knowledge model based on Category II ½ curves assessed most of the reach with high suitability, but the shores and the deep areas corresponding to the northern and middle reach areas (Fig. 11 A). On the contrary, the Expert-knowledge model based on the Category III HSCs presented most of the area as medium suitability and higher variability; the suitability gradually increased from the shallower areas to the deeper, where it can be High or Very High (Fig. 11 B).

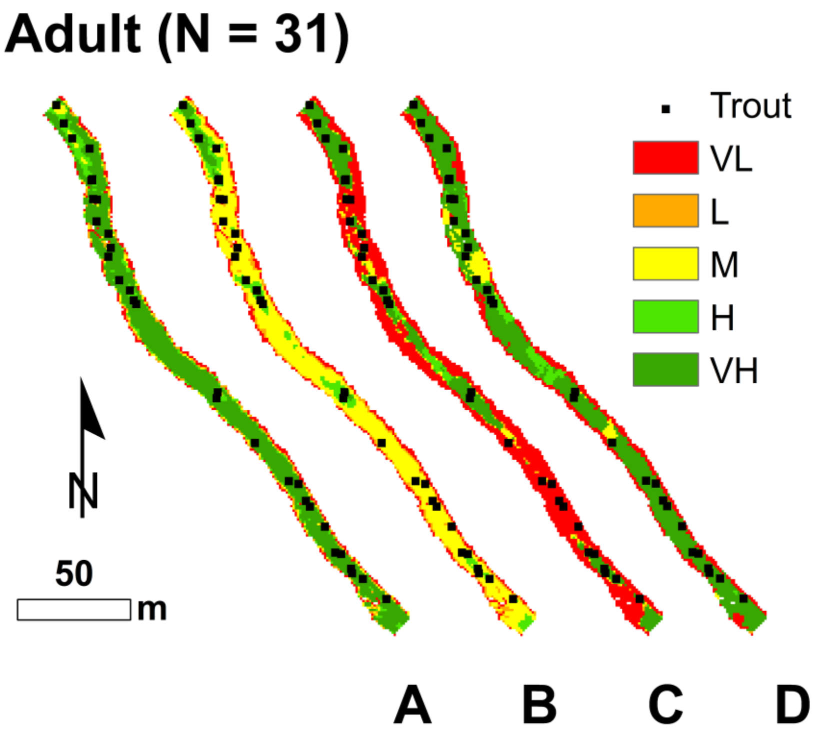

Fig. 11. Maps of habitat suitability assessment for the adult brown trout carried out with the four models in the validation site (Cabriel River). A - Expert-knowledge based on Category II ½ curves, B - Expert-knowledge based on Category III curves, C - Unmodified Data-driven model and D - Modified Data-driven model. N means trout observations at the validation site. The Suitability was classified in 5 categories corresponding to Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH).

The unmodified Data-driven model assessed most of the reach as unsuitable for adult brown trout, but the deep areas which were assessed with the highest suitability (Fig. 11 C). One rule was modified in order to maximize the number of fish locations assessed with the maximum suitability (Rule #14, Table 7). The assessment showed most of the reach as highly suitable but the shores remained unsuitable (Fig. 11 D).

4.5.2 Fuzzy models performance - Juvenile

In general, there was a larger agreement among the models for the juvenile brown trout. The Expert-knowledge fuzzy model based on Category II ½ curves assessed most of the reach with the highest suitability, but the shores and the deeper areas, especially in the northern area (Fig. 12 A). The Expert-knowledge model based on Category III HSCs presented the highest suitability all along the surface but a narrow fringe parallel to the shores (Fig. 12 B).

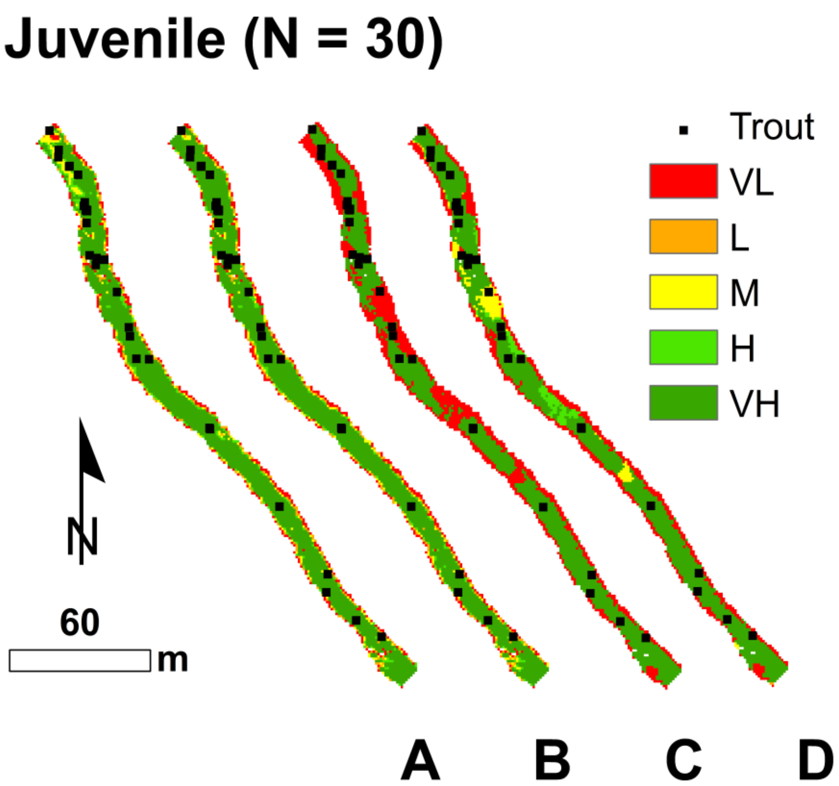

Fig. 12 Maps of habitat suitability assessment for the juvenile brown trout carried out with the four models in the validation site (Cabriel River). A - Expert-knowledge based on Category II ½ curves, B - Expert-knowledge based on Category III curves, C - Unmodified Data-driven model and D - Modified Data-driven model. N means trout observations at the validation site. The Suitability was classified in 5 categories corresponding to Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH).

The unmodified Data-driven model assessed most of the reach as highly suitable but a relatively wider fringe parallel to the shores (Fig. 12 C) and small pieces of shallow/fast habitats. One rule was modified in order to maximize the number of fish location assessed with the maximum suitability (Rule #4,Table 7). This change improved the sensitivity of the model, and the validation flow was assessed mostly with the highest suitability, enlarging the suitable area and narrowing the aforementioned fringes (Fig. 12 D). As the previous Data-driven model, this one presented small pieces of shallow/fast habitats with lower suitability.

4.5.3 Fuzzy models performance - Fry

The habitat assessment for the fry showed certain disparity between models. The Expert-knowledge model based on Category II ½ curves assessed most of the reach as highly suitable but the shores and the deeper areas (Fig. 13 A); however, the Expert-knowledge model based on Category III presented high suitability all along the surface but keeping certain decrease on the deeper areas, although this phenomenon appeared attenuated in comparison with the previous model (Fig. 13 B).

On the contrary, the unmodified Data-driven fuzzy model assessed the deeper areas as highly suitable and the shallower as unsuitable, switching the habitat assessment of the two Data-driven models (Fig. 13 C). One rule was modified to improve the Data-driven fuzzy model (Rule #15, Table 7), maximizing the number of fry locations assessed with the maximum suitability; thus the reach was assessed mostly with the highest suitability but near the shores (Fig. 13 D), and the difference with the previous Data-driven model was notable.

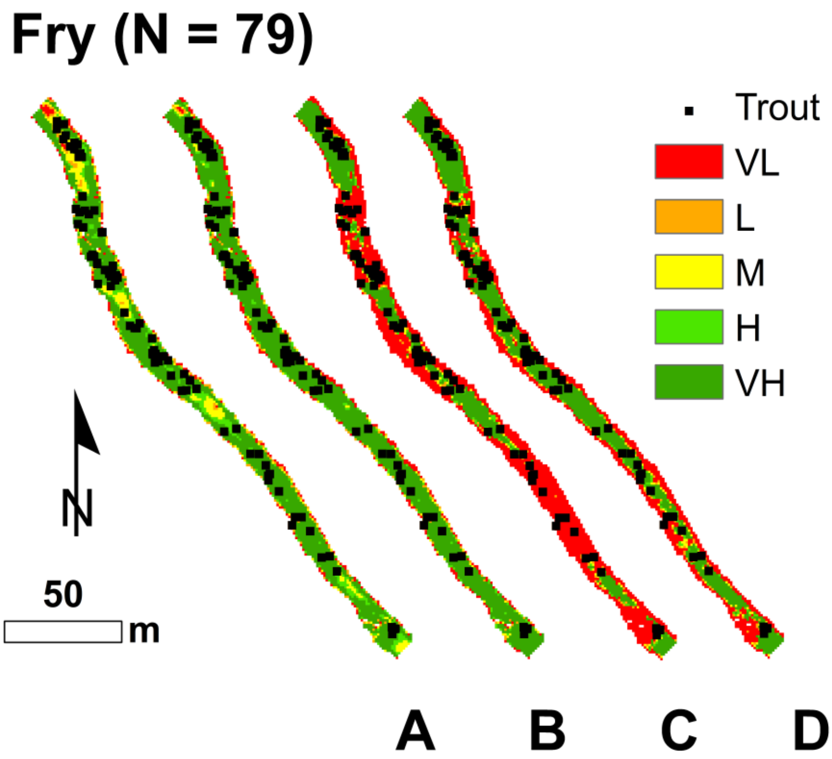

Fig. 13 Maps of habitat suitability assessment for the fry brown trout carried out with the four models in the validation site (Cabriel River). A - Expert-knowledge based on Category II ½ curves, B - Expert-knowledge based on Category III curves, C - Unmodified Data-driven model and D - Modified Data-driven model. N means trout observations at the validation site. The Suitability was classified in 5 categories corresponding to Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH).

4.6 Comparison of Fuzzy models performance - Frequency analysis

The frequency analysis of the habitat assessment over the simulated validation site showed that, disregarding the Data-driven modifications, the Expert-knowledge models were in general more optimistic (higher suitability values) than the Data-driven models (Fig. 14).

4.6.1 Expert-knowledge fuzzy models based on Category II ½ HSCs

The Expert-knowledge fuzzy models based on Category II ½ curves indicated, for the three size classes, that most of the reach had the maximum suitability; thus the sensitivity was high but not the specificity (Fig. 14, upper sequence). The models presented large frequency of the higher suitability (i.e. suitability ranging from 0.6 to 1.0) within the entire reach (i.e. Availability) (Fig. 14 black bars) similarly than the frequency analysis of the assessment of the trout locations (i.e. habitat Use) (Fig. 14 grey bars).

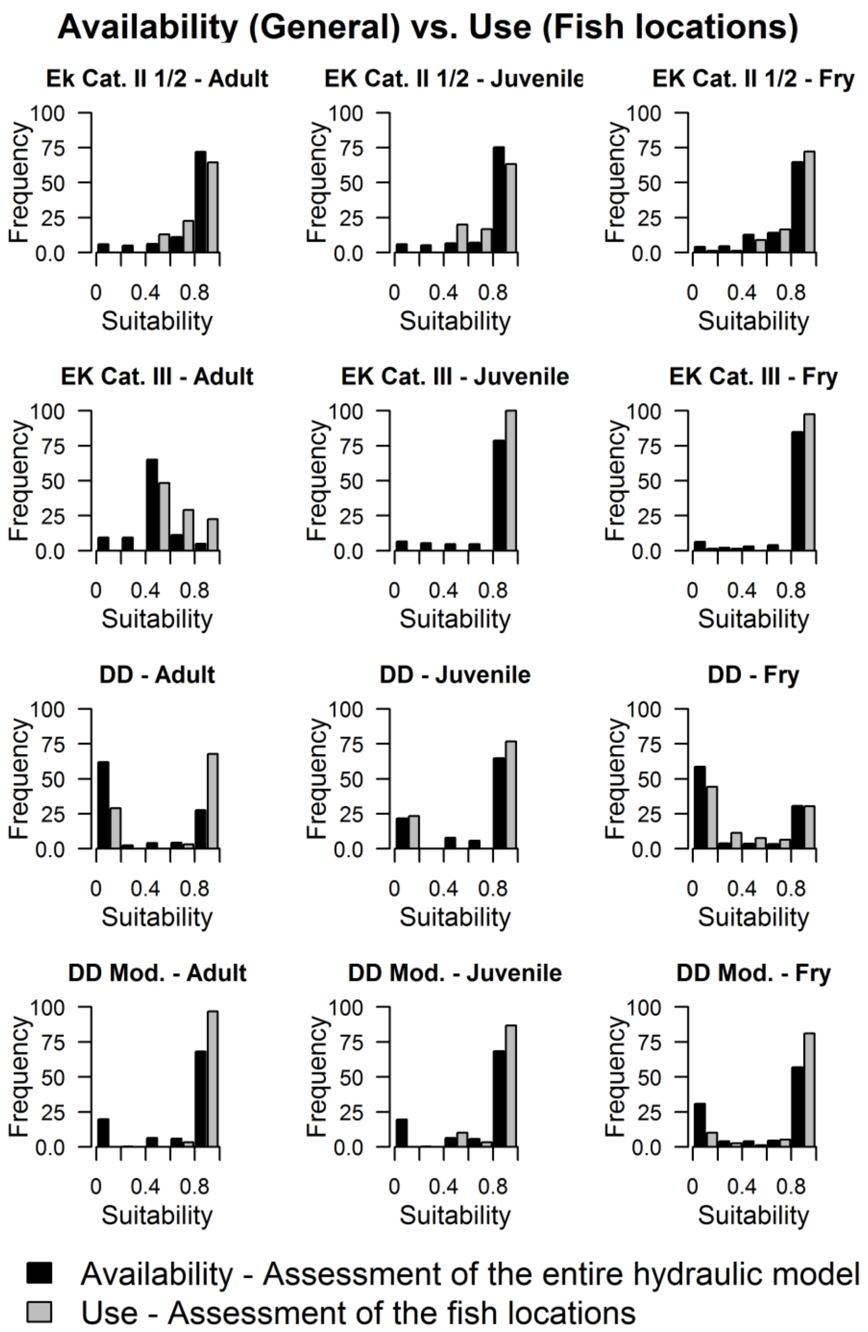

Fig. 14 Frequency analysis of the habitat assessment carried out with the four generated models and the three size classes (black bars) over the entire simulation reach (Availability data). Frequency analysis of the habitat assessment carried out with the four generated models and the three size classes (grey bars) over the corresponding size class locations (Use data). EK mean Expert-knowledge and DD Data-driven.

4.6.2 Expert-knowledge fuzzy models based on Category III HSCs

The frequency analysis for the Expert-knowledge fuzzy models derived from Category III HSCs (Fig. 14 second sequence) presented most of the area with the maximum suitability (i.e. 0.8 - 1.0) for the juvenile and the fry size classes (Fig. 14 black bars) and all the observed trout locations in both size classes were assessed with the maximum suitability (Fig. 14 grey bars). In that class of maximum suitability, there is an over-proportion of habitat Use in relation to Availability. Therefore, the models for juvenile and fry based on Category III HSCs presented perfect sensitivity but low specificity.

The adult model presented the habitat assessment spread along the considered categories in comparison with the previous models. The largest frequency of the Availability (Fig. 14 black bars) appeared for the middle ranged suitability (i.e. suitability from 0.4 to 0.6) and the frequency decreased in both sides, towards lower and higher suitability (Fig. 14 black bars). However, there were no habitat Use in the lower suitability intervals (Fig. 14 grey bars) and an appreciable amount of trout locations were assessed within the higher suitability intervals (i.e. suitability ranging from 0.4 to 1.0). As well as in the two previous models, there was an over-proportion of habitat Use in relation to the Availability; this means that there was a positive selection of the fish towards the microhabitats of higher suitability, as expected. Therefore it presented good sensitivity but the better specificity among the Expert-knowledge fuzzy models (Fig. 14 first and second sequences).

4.6.3 Data-driven fuzzy models

The unmodified Data-driven fuzzy models presented major disparity among the results (Fig. 14 third sequence). The adult case presented a good trade-off between sensitivity and specificity (Fig. 14 third sequence - Adult). The largest frequency of the assessment of the entire reach appeared in the lowest suitability interval (i.e. suitability ranging from 0.0 to 0.2) (Fig. 14 black bars) whereas the maximum frequency of adult locations corresponded to the highest suitability interval (i.e. 0.8 - 1.0) (Fig. 14 grey bars). However, accordingly to the premise that overprediction is not necessarily an ecological error, the Data-driven model for the adult class was modified by enlarging the adult locations assessed with the maximum suitability (Fig. 14 last sequence - Adult). It maximized the trout locations assessed with the maximum suitability (i.e. suitability ranging from 0.8 to 1.0) (Fig. 14 grey bars) thus maximized the sensitivity and retained certain specificity, because the Availability data presented a relevant proportion with the lowest suitability (i.e. suitability ranging from 0.0 to 0.2) (Fig. 14 black bars).

The frequency analysis of the unmodified Data-driven model for juvenile showed most of the reach and most of the habitat use at microhabitats with the highest suitability (i.e. suitability ranging from 0.8 to 1.0) (Fig. 14 third sequence - Juvenile). Thus presented a high sensitivity and relatively low specificity (Fig. 14). However an appreciable amount of trout were located in areas assessed as unsuitable (i.e. suitability 0.0 - 0.2). The modification of a single rule (Rule #4 Table 7) displaced the assessment (Fig. 14 last sequence - Juvenile) of that juvenile locations to higher suitable values (Fig. 14 grey bars) but the specificity remained almost constant (Fig. 14 black bars) .

The unmodified Data-driven fuzzy model for the fry size class presented the worse results (Fig. 14 third sequence - Fry). The maximum frequency appeared for the lowest suitability (i.e. suitability 0.0 - 0.2) and most of the fish locations were assessed as unsuitable (Fig. 14 grey bars). Thus the sensitivity was low. The modification of a single rule (Rule #15 Table 7) improved the results (Fig. 14 last sequence - Fry) maximizing the trout locations with maximum suitability (i.e. 0.8 - 1.8) (Fig. 14 grey bars), but keeping an acceptable trade-off between sensitivity and specificity.

4.7 Comparison of Fuzzy models performance - Fuzzy Kappa analysis

The spatially explicit comparison between models did not show a clear pattern and the degree of similarity varied regarding the considered approach and size class. However it could be considered that the similarity was generally low.

The adult models presented no similarity regardless the considered approach or model (Table 8) but a slight similarity between the Expert-knowledge fuzzy model based on Category II ½ HSCs and the modified Data-driven model, because this comparison achieved a value of fuzzy Kappa 0.42. However this value was mainly produced by the fact that most of the reach was considered highly suitable rather than the fact that both approaches presented similar pattern of the assessment; more specifically, the Expert-knowledge approach indicated a reduction of suitability as the depth increases whereas the unmodified Data-driven model presented the opposite pattern (Fig. 11).

The juvenile models presented perfect agreement between those assessments belonging to the same approach, but lower similarity for different approaches, although the modified Data-driven model presented certain similarity with the Expert-knowledge models, achieving the fuzzy Kappa values of 0.42 and 0.5 in comparison with the models based on Category II ½ and Category III curves, respectively (Table 8).

The fry models presented low similarity between the models derived under the same approach. Besides there was no similarity between the models with different approaches. It was remarkable that the agreement expected by chance was larger than the expected by any coincidence on the model assessment, showing negative values of the fuzzy Kappa in the comparison of the Expert-knowledge fuzzy model based on Category II ½ and its unmodified Data-driven counterpart (Table 8). Again, the results confirm the relevant differences between the unmodified Data-driven model and any of the others.

Table 8. Spatially-based comparison among the models for each size class. The values correspond to the fuzzy Kappa statistic (Hagen, 2003). The habitat assessment was reclassified in 5 equal-length intervals, the degree of membership was linearly decreasing from perfect agreement (1 to the same category) to null agreement (0 to the further category). The radius of influence was 5 m (adult and juvenile) and 2.5 m (fry).

		EK Cat II ½	EK Cat III	Unmodified DD	Modified DD
Adult	EK Cat II ½	1	0.2	0.12	0.42
	EK Cat III	-	1	0.3	0.26
	Unmodified DD	-	-	1	0.3
	Modified DD	-	-	-	1
Juvenile	EK Cat II ½	1	0.87	0.33	0.43
	EK Cat III	-	1	0.35	0.5
	Unmodified DD	-	-	1	0.76
	Modified DD	-	-	-	1
Fry	EK Cat II ½	1	0.52	-0.07	0.12
	EK Cat III	-	1	0.09	0.24
	Unmodified DD	-	-	1	0.51
	Modified DD	-	-	-	1

4.8 Fuzzy models performance - Density and suitability correlation analysis

4.8.1 Density analysis

The density analysis showed that the adult brown trout appeared sparsely distributed all along the reach but with a peak of density in the northern part of the validation site (Fig. 15). Although the juvenile brown trout appeared all along the study site (Fig. 15) a density peak appeared close to the adults, which conditioned the generated density categories (Fig. 15). Both cases presented the maximum density in areas with a relatively large depth and in an area where the flow was concentrated downstream a relatively fast habitat. The fry appeared more sparsely distributed with several areas of Very High density; in general they were far from the older individuals (Fig. 15).

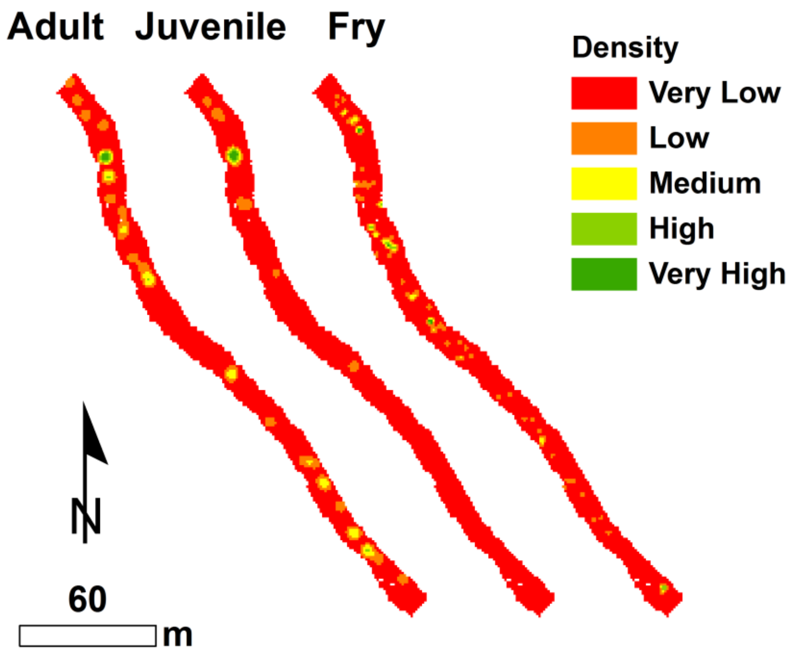

Fig. 15 Density map. The kernel density tool in ESRI® ArcMapTM 9.3 (Copyright© 1999-2008 Esri Inc.) was used to calculate densities and the results were standardized between zero and one and divided in five equal-length intervals corresponding to Very Low, Low, Medium, High, Very High density.

4.8.2 Density and suitability correlation - Adult

The analysis of correlation of the assessed suitability and the observed density showed different patterns regarding the methodology and the size class, without a common pattern (Fig. 16). The Expert-knowledge model based on Category II ½ HSCs for adult showed a positive trend between the average suitability and the density, but a decrease on the average suitability in the most densely populated interval (Fig. 16 upper plot). The Expert-knowledge fuzzy model based on Category III HSCs presented a positive correlation between the density and the average suitability achieving the maximum suitability in the most highly populated areas (Fig. 16 upper plot). The unmodified Data-driven presented an increasing trend as the density increases, whereas the modified Data-driven presented similar pattern but the average values were higher than in the unmodified counterpart.

4.8.3 Density and suitability correlation - Juvenile

The Expert-knowledge fuzzy model based on Category II ½ HSCs for the juvenile class presented a flat trend as the density increases from Very Low density to Medium density, and decreased for larger densities. The most densely populated areas dit not assess any area with the highest suitability (Fig. 16 middle plot). The Expert-knowledge fuzzy model based on Category III HSCs presented high average suitability for the lowest density interval achieving the maximum suitability for the remaining density intervals regardless the observed density interval. The unmodified Data-driven fuzzy model for the juvenile class presented medium average suitability with an slight increasing trend as the density increases, although it presented some irregularities (Fig. 16 middle plot). The modified Data-driven fuzzy model presented an increment of the average suitability at the lower density intervals presenting the maximum average suitability for the remaining intervals.

4.8.4 Density and suitability correlation - Fry

The Expert-knowledge fuzzy models based on Category II ½ and in the Category III HSCs presented high average suitability regardless the density interval (Fig. 16 lower plot). The unmodified Data-driven fuzzy model presented positive correlation between the average suitability and the fry density (Fig. 16 lower plot). However, as the juvenile, it presented some irregularities. The modified Data-driven fuzzy model presented and improvement regarding the unmodified Data-driven fuzzy model with higher values of the average suitability at any considered density interval, and not presenting pronounced irregularities.

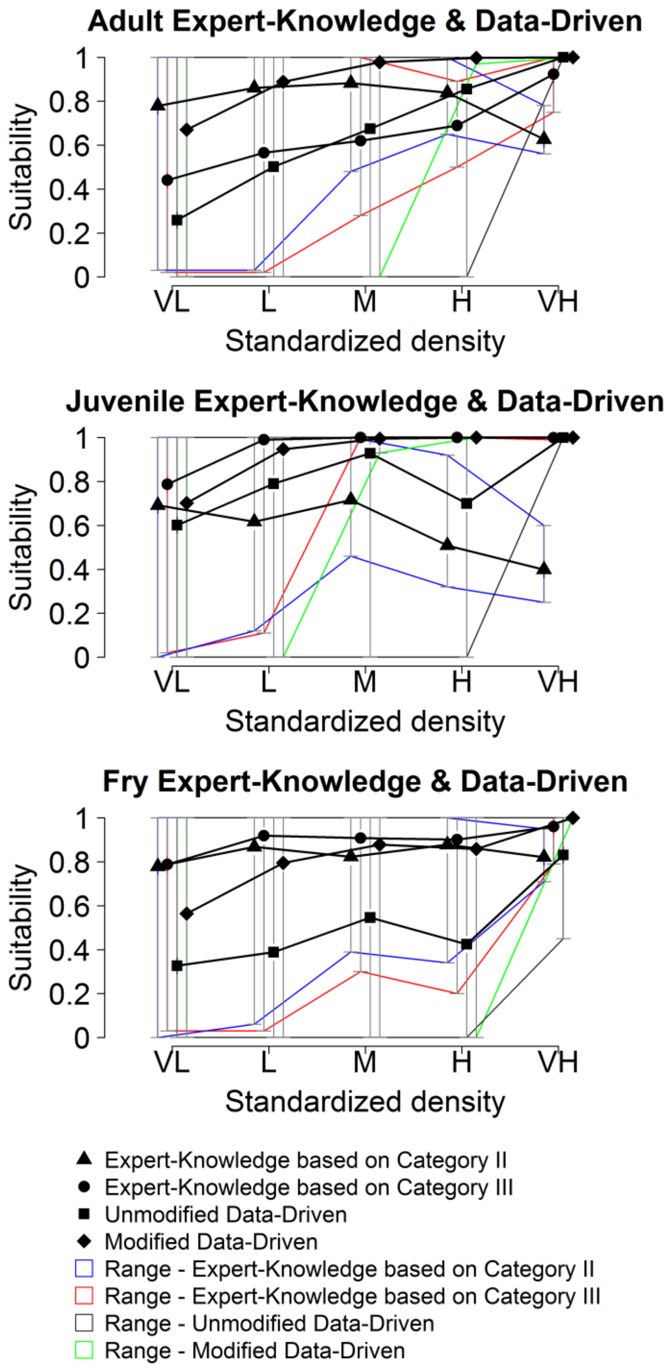

Fig. 16. Average suitability (black lines) for the five generated density categories, by fish size class. The frequency distribution of the assessed suitability on each density area was very skewed, thus the range of the habitat assessment per density category were plotted. These values corresponded to the maximum and the minimum values assessed in the cosidered area.

5 Discussion

As general overview about the results, they showed that some of the habitat suitability models are appropriate to be applied in the Cabriel River at the validation site or even further, because they provided with satisfactory results in terms of performance in both terms, in number of correctly assessed fish locations and taking into account an assessment of fish density.

5.1 Habitat Suitability Curves - HSCs

5.1.1 HSCs Category II ½

The development of Category II ½ suitability curves followed common standards (Bovee, 1986). These curves differed of those from literature in most of the cases.

The adult curves allowed us the greatest number of possible comparisons, as explained in this paragraph. In general the Category II ½ curves presented here showed the highest suitability for low values of both velocity and depth, compared with some of the most relevant studies about adult brown trout. Such reference studies included curves of the three Categories, I, II and III (Bovee, 1978; Raleigh, 1984; Hayes and Jowett, 1994; Heggenes, 1996; Ayllón *et al.*, 2010) (The last study not as important than the previous, but recently developed in the Iberian Peninsula). Our curve of velocity was similar to those from Bovee and Raleigh, and the depth curve in comparison with the Ayllón's curve and the corresponding from Hayes and Jowett. Notice that Hayes and Jowett's study presented both, curves of Category II and Category III. The substrate showed similar suitability over coarse substrates than those from literature, and did not present remarkable differences.

For juvenile trout, the Category II ½ curves presented less possibilities for comparison. The velocity curve presented similar pattern than Raleigh's and Bovee's, but Ayllón's differed strongly because it showed the highest suitability for fast waters. The depth curve presented the higher suitability for shallower areas compared with Raleigh's; regarding that from Bovee, it entirely covers our Category II ½ curve, becoming wider, whereas the Ayllón's presents higher suitability for deeper areas than our curves. The substrate suitability present similar pattern than those from literature with the highest suitability over coarse substrates.

The Category II ½ curves for fry presented the highest suitability for lower velocities, compared with Bovee's, Raleigh's and Ayllón's curves. Nevertheless, the depth curve was similar to Ayllón's, lower than Raleigh's; like the previous case, Bovee's curve (very generalist, obtained from different sources) covers entirely our curve. A great discrepancy appears with the substrate, because the aforementioned studies which include any reference to fry showed the highest suitability similarly to the older size classes, i.e. over coarse substrates. In contrast, our curve presented the maximum for silt. This result could be strongly determined by a limited availability of the combinations depth-velocity-substrate, with an important correlation of habitats, meaning that the fry usually occupy the slow and shallow habitats, which in turn are usually covered with silt; this curve could produce a bias in the physical habitat simulation, thus any application should be very careful in this aspect, considering the curves from other authors additionally.

5.1.2 HSCs Category III

The Category III curves developed in the present study showed relatively major similarities with those from literature than the Category II ½.

The adult case presented the highest suitability for faster areas in comparison with Bovee's, Raleigh's and Ayllón's curves, but similar if compared with the corresponding of Hayes and Jowett. The curve for depth had the same pattern of similarities, the highest suitability was presented for deeper areas in comparison with all of the curves from literature but the Hayes and Jowett's and the Ayllón's. The later presents a constant high suitability regardless the depth in accordance to Bovee's curves (Bovee, 1978).

The juvenile presented the curves with the largest coincidence with those from literature; they were different from Raleigh's and Bovee's, presenting the highest suitability for faster flows but with similar pattern than the Ayllón's. In addition, the curve for depth showed clear similarities with Bovee's, Raleigh's and Ayllón's. The substrate did not differed substantially, and the highest suitability was just slightly displaced one category toward the finer substrates in comparison with the corresponding Category II ½ curve, but with an appreciable suitability at coarse substrates (cobble/boulder), like the curves from literature.

The fry curves presented similar pattern than Ayllón's and Raleigh's for velocity, whereas the Bovee's curve involves the fry's Category III curve for the variable velocity. The depth curve presented similar pattern than Bovee's and Raleigh's but the highest suitability for deeper areas if compared with Ayllón's. The displacement produced by the forage ratio for the variable substrate did not generate strong changes in the Category III curve (compared with the Category II ½) thus the curve presented the highest suitability also over silt, differing from those from literature which present the highest suitability over coarse substrates.

5.2 Expert-knowledge fuzzy models

5.2.1 Expert-knowledge fuzzy models based on Category II ½

The fuzzy models derived from the Category II ½ HSCs presented good performance in terms of assessment of the fish locations for the three size classes; the fish locations were assessed mostly over 0.8 suitability. Although the specificity was low, most of the area was also assessed over 0.8 suitability, but it was not considered misleading since the study reach homes a stable trout population (Martínez-Capel *et al.*, 2009a) and over-prediction should not be always considered as a model error (Mouton *et al.*, 2010b).

The limited amount of trout and the extreme variability in natural systems are factors limiting the colonization of all the suitable microhabitats thus this uncolonised microhabitats that presented similar habitat conditions than the occupied one's should, likewise, be considered suitable. Therefore these models become numerically overpredictive. However, the patterns of habitat suitability over the study reach showed a decay of the suitability over the deeper areas (Fig. 16). This fact was especially significant for the adult and the juvenile size classes, because it coincided with their most densely populated areas (Fig. 11, Fig. 12 and Fig. 15). The current flow during the time of the validation survey (Q = 0.89 m^3/s) represented the flow Q_{85} of the stream flow time series; thus larger depths are expected at the study site, and the validity of the models based on Category II ½ curves would be easily overrode.

Several authors considered that, for the large trout, the large depth could not be considered as limiting at all (Bovee, 1978; Ayllón *et al.*, 2010) and specifically some of them pointed out the fact that as the trout ages it becomes more and more pool dweller

(Heggenes, 1996; Ayllón *et al.*, 2010). These asseverations agreed with previous studies where adult and juvenile trout were repeatedly observed in deeper areas in the vicinity of the validation site (Martínez-Capel *et al.*, 2009a). The trout is considered a territorial fish (Chapman and Bjonn, 1969; Titus, 1990; Johnsson *et al.*, 2000) and this territoriality is related with food availability (Brännäs et al., 2003) but not related with food scarcity.

Trout has been demonstrated to be a drift-feeding strategist (Elliott, 1973; Bachman, 1984) holding stations in slow water, but close to a fast current (Wańkowski and Thorpe, 1979; Bachman, 1984) and recently some models included the availability of macroinvertebrates-drift to improve fish habitat modelling with promising results (Hauer *et al.*, 2012) agreeing with the previous asseverations. These studies suggests that our observations of trout distribution at the validation site matches the food availability pattern. An implication for future studies is the possibility of improving the suitability for trout in that deeper and more densely populated areas, but keeping a good trade-off between sensitivity and specificity.

5.2.2 Expert-knowledge fuzzy models based on Category III

The calculation of Category III HSCs was made in order to compare and evaluate their application in the development of Expert-knowledge fuzzy models. Given the previous comments about the habitat assessment in deep areas, the Category III curves meant a possibility for improvement, due to the shift of curves produced by the forage ratio.

The generated Category III curves presented all the inconveniences compiled by Payne (2009). The application of the forage ratio produced a displacement of the curve toward the higher values of the variable in the case of velocity and depth, whereas the substrate curves remained almost constant. In the past, this displacement (over-correction) as well as the statistical assumptions behind the forage ratio, discouraged their application (Bovee, 1996).

However, upholding their consideration, some authors have demonstrated habitat availability affects habitat use and even habitat selection by brown trout (Heggenes, 1991; Rincon and Lobon-Cervia, 1993; Grossman and De Sostoa, 1994) recommending certain consideration or correction based on habitat availability. Although the overcorrection produced through the application of the forage ratio is undesired, and it should be prevented with certain quality control during the generation and processing of data (Payne, 2009), and despite the existence of several alternatives, they are not widespread. Hayes and Jowett (1994) pointed out that the forage ratio (i.e. used and availability proportions) is particularly sensitive to extreme values, and it does not account for habitat that was not available at the time or place of sampling, therefore the tails of the curves should not be considered.

Regardless this important considerations, here the generated curves were used just as a base to develop the Expert-knowledge fuzzy models and the tails of the curves were not trimmed. The Category III HSCs agreed with previous studies where the preference curves for adults shifted to deeper and faster areas compared with suitability ones (Bovee and Zuboy, 1988).

In contrast, the curves for fry presented similar displacement, whereas the literature suggested that the forage ratio could displace the curves to shallower and slower areas.

The assessments with the Expert-knowledge models based on Category III curves were considered in general satisfactory; however, this result cannot be considered as general for any study and the considerations by previous authors are very important for any application (e.g. Bovee, 1996; Payne, 2009). The frequency analysis on habitat availability and fish habitat use showed that such models improved the models performance (Fig. 14). The specificity was reduced in some caes but in contrast they presented clearer positive correlation between the average suitability and the trout density (Fig. 16).

The results for adult brown trout could be criticized because most of the trout location were assessed as Medium, whereas the models based on Category II ½ indicated mostly the maximum suitability at those locations. It should be noticed that suitability 0.4 - 0.6 did not mean absence, and this range of suitability corresponded to the intervals of the curves which presented certain suitability, thus trout were observed but they were not abundant. Therefore, it was considered a satisfactory result.

The model for adult trout can be considered the most specific among the models presented here, because the trout covers more intensively the areas with Medium-to-High suitability than the remaining models (Fig. 11), and the density and average suitability was clearly correlated with an increasing trend (Fig. 16). The Juvenile was the least specific and most of the area was assessed with the maximum suitability (Fig. 9); nevertheless, the least densely populated areas presented the lowest average suitability. The model generated for the fry size class presented only a small improvement; but the model based on Category II ½ presented a slight decrease at the more densely populated areas, while the other model based on the Category III HSCs presented a slight increase.

5.3 Data-driven fuzzy models

Regarding the Data-driven fuzzy models the achieved values of the Cohen's Kappa could be considered acceptable in comparison with previous studies which used similar training strategies (Mouton, 2008; Muñoz-Mas *et al.*, 2012). Considering the amount of trout locations at highly suitable microhabitats, the Data-driven fuzzy models (unmodified) showed a poorer performance than the Expert-knowledge approach (Fig. 14) but they presented a positive correlation between the assessed suitability and the trout density in any case (Fig. 16). The observed deficiencies were improved by modifying the corresponding rules thus providing finally satisfactory results.

5.3.1 Data-driven fuzzy model - adult

The Data-driven fuzzy model for the adult size class was considered satisfactory despite the relatively low value of the performance criteria (Kappa = 0.31). The model presented the best trade-off between sensitivity and specificity (Fig. 14) in addition to a positive trend between average suitability and trout density (Fig. 16). These results agreed with previous studies (Heggenes, 1996; Ayllón *et al.*, 2010) which assigned the maximum suitability to the deepest areas. The modification of this Data-driven fuzzy model was mainly considered in the development of the proposed methodology. In order to maximize the model agreement one rule was modified (Rule #14 Table 7). The modified Data-driven model improved the habitat assessment over the trout locations, displacing all of them to the interval comprising the maximum suitability but keeping certain specificity (Fig. 14), thus the model was considered suitable for further analysis. However, the rule that includes the most extreme condition -High velocity, High depth, High substrate- was determined as High. This will provide an increasing suitability as the flow increases, and it should be taken into account if this model is applied in further analysis; our results indicate that the limiting conditions for the adult brown trout had not been surveyed. One of the advantages of the fuzzy approach is the versatility and easy adjustment, thus this deficiency could be fixed adding an extra Fuzzy Set to the variable depth or velocity and determining the rule consequent in those new cases as Low.

5.3.2 Data-driven fuzzy model - juvenile

The Data-driven model for the juvenile class was also considered satisfactory despite it achieved the lowest Kappa value (Kappa = 0.21). The frequency analysis over the validation site and the assessment at the trout locations presented a acceptable trade-off between the sensitivity and the specificity (Fig. 14). The modification of this model implied the modification of a single rule (Rule #4 Table 7), displacing the habitat assessment over the trout locations close to the maximum suitability for all the observed individuals; the frequencies of the available unsuitable pixels remained almost unaltered (Fig. 14).

The analysis of the consequents showed certain discrepancy with the literature because usually it has been considered that the habitat suitability decreases beyond certain depth but the Data-driven model determined the rules' consequent including the High depth as High (Bovee, 1978; Raleigh, 1984; Ayllón *et al.*, 2010). However, in contrast to the adult model, the Fuzzy Rules including the maximum velocity were assessed as Low in any case; this issue allow the application of the model in larger flows. On the other hand, a careful application is necessary in the habitat assessments when considering Low velocity, Low dept and High substrate, since it has been assessed as Highly suitable and could provide the maximum suitability over too shallow areas. In that sense, a narrow Fuzzy Set covering the extremely shallow depth would be preferable in order to avoid the aforementioned situation.

5.3.3 Data-driven fuzzy model - fry

The data-driven models for the fry class were clearly the worst, despite the highest value of the performance criteria (Kappa = 0.37). The frequency analysis in validation determined the maximum frequency of fish in the unsuitable microhabitats. This results were not surprising because the database for fry had the lowest sample size (N = 44) which hardly could cover the whole conditions found in the study site, even taking into account that the validation sample was almost twice as large (N = 79). The fry model considered that small fish prefer pools in accordance to previous studies but disagreeing with the distribution pattern in the spatial validation (Fig. 15). The observed pattern (Fig. 15), accordingly with several authors, could be produced by the exclusion produced by the presence of older and larger trout, which also prefer these areas (Raleigh, 1984). However, the Expert-knowledge model did not show that pattern, thus this conclusions could come from imperfections on the database or properly from the exclusion of the better habitats derived from the presence of older individuals. Therefore further effort should be placed in the improvement of the fry's database in order to discern the causes of that phenomenon.

5.4 Rules comparison

The comparison of the Fuzzy Rules developed with Expert-knowledge approach and the rules corresponding to those Fuzzy Sets but trained with a Data-driven approach showed that the Expert-knowledge approach was more optimistic, i.e., it gave higher suitability in general. This result was not surprising because the consequent of the

Data-driven model was presence/absence and the training strategy of the Data-driven approach was based on the principle of the-winner-takes-all (i.e. the most frequent result in the binary result was the selected consequent, High or Low). In contrast, the Expert-knowledge approach presented a smoother transition, considering as Medium those rules that presented certain suitability on the combination of the HSCs. The results also pointed out the improvement of the model performance in Data-driven models when Fuzzy Sets are developed based on the calculation of the Shannon-Weaver entropy, in accordance with previous studies (Mouton, 2008); thus these Fuzzy Rules sets presented lower values of the Cohen's Kappa and larger amount of uncovered rules.

5.4.1 Rules comparison - adult

The models based on Category II ½ curves for adult trout presented two discrepant rules. The first one corresponding to, Low Velocity, Medium Depth and Low Substrate (Rule #3, Table 7). This rule comprised 30 cases and the Expert-knowledge approach determined it as High whereas the Data-driven approach as Low. The second discrepant rule corresponded to: High Velocity, High Depth and High Substrate (Rule #24, Table 7) but trained with a single datum. In both cases the discrepancy agreed with the 'pure' Data-driven approach that relegate the adults to deep areas in accordance with several studies (Ayllón *et al.*, 2010) that demonstrated the preference for pools of the adult brown trout.

The comparison of the rules based on the Category III HSCs presented only one discrepant rule: Medium Velocity, Low Depth, High Substrate (Rule #18, Table 7). The Expert-knowledge approach determined it as High whereas the Data-driven approach as Low. Taking into account the fact that the Medium depth in the Category II ½ case corresponds mostly to Low depth in the Category III case, this conflict also arose because the Data-driven approach considered the shallow areas as unsuitable in any case, while the Expert-knowledge approach was more permissive and determined some of them with better suitability.

Altogether suggested a review of the rules including the shallow to medium depth conditions in order to carry out a downwards adjustment and the opposite for the rules that include the large depth. Additionally the importance of the substrate should be also reviewed since some of that optimistic consequents were conditioned by its partial suitability, for instance, the Rule #18 (Table 7).

5.4.2 Rules comparison - juvenile

The models based on Category II ½ curves presented one discrepant rule corresponding to: Low Velocity, Low Depth and Low Substrate; it was trained with 15 cases. The Data-driven approach indicated Low suitability for Low velocity and Low substrate regardless the depth, while the Expert-knowledge approach considered some combination as Medium or High. The Data-driven output was in accordance with previous studies (Jutila *et al.*, 1999) which demonstrated, applying a multivariate approach, the positive correlation between depth and substrate size with juvenile density. According to the present results, these rules (including Low velocity - Low substrate) had a slight impact on the assessment of the validation site and no modification was considered here. However, in some cases such rules should be adjusted in the Expert-knowledge model, for the application in different rivers with other physical characteristics.

The models based on Category III presented one discrepant rule, that is: Low Velocity, Medium Depth and High Substrate (Rule #15, Table 7), trained with a relatively large amount of data (44 cases) corresponding to the 16 % of the whole database. The Expert-knowledge assessed that rules consequent as High but the robustness of the rule could suggest a slight reduction on the rule consequent moving from High to Medium approximating the result to the obtained by means of the Data-driven approach. However, regardless the great differences in the inputs discretization (i.e. differences in the Membership Functions associated to the Fuzzy Sets) those conditions were assessed also as highly suitable in the Data-driven model, with an improvement of model performance. The juvenile brown trout has been reported to inhabit at shallower water than adults with a relatively slow velocity and coarse substrate (Raleigh, 1984), which suggest to disregard any modification.

5.4.3 Rules comparison - fry

The models based on Category II ½ for brown trout fry presented two discrepant rules; the first was, Low Velocity, Low Depth, High Substrate (Rule #14, Table 7). The Expert-knowledge approach assessed the suitability as High, whereas the Data-driven approach did the opposite. The second discrepant rule was, High Velocity, Low Depth, High Substrate (Rule #22, Table 7); in this case the Expert-knowledge indicated Low and the Data-driven High. The first discrepant rule was trained with 37 cases, corresponding to the 42 % of the data, whereas the second corresponded to the 8 %. Both rules suggested a preference for deep areas instead the observed distributions at the validation site (Fig. 13)

The models derived from Category III curves also had two discrepant rules; considering the smaller amount of Fuzzy Sets determined for depth ('Very Low' depth was no considered after the Category III curves), these discrepant rules corresponded to the previous ones observed after the Category II ½-based models. The habitat evaluation in the validation site with the Data-driven model (Fig. 14) and the analysis of correlation with density (Fig. 16) did not suggest any clear or robust modification of the Expert-knowledge model based on the results obtained here because the sample size was larger in the validation database and the Expert-knowledge outperformed the Data-driven models. However, the Data-driven approach apparently corroborated previous studies where the fry preference for deeper areas was considered but the presence of older individuals would displace them toward shallower areas (Raleigh, 1984). Therefore further research should be placed on the enlargement of the fry database and on the study of the real habitat suitability for the fry class.

5.5 Comparison based on Fuzzy Kappa

The spatially explicit comparison among the models corroborated certain similarity in each approach, especially in the cases where no relevant displacements of the HSCs were observed. Obviously, regarding the Data-driven models a great concordance was expected because, depending on the model, only one or two rules were adjusted or modified. The spatial analyses showed major similarities in juveniles, followed by adults and fry.

The highest values of suitability were provided by the juvenile's models; however, most of the area was assessed with the maximum suitability, which limited the achievement of any conclusion about the properness of the entire analysis. These models presented almost perfect agreement between pairs and slight similarities between approaches.

On the other hand, the values achieved in the adult's models comparison disagreed with the expected results. The model based on Category III HSCs apparently provided the highest suitability in similar areas than the unmodified Data-driven (Fig. 11); the major discrepancies appeared in the shallower areas where the Expert-knowledge approach had been more permissive providing medium suitability, in comparison with the Data-driven (Table 7). In contrast, the fuzzy Kappa was relatively low when comparing Expert-knowledge fuzzy model based on Category III HSCs with the unmodified Data-driven fuzzy model (Table 8), and similar in magnitude as the comparison of the unmodified Data-driven fuzzy model with the Category II ½ based model (Table 8). Therefore, the comparison did no shed any light on adult models' similarity.

The fry models presented relatively low values of the fuzzy Kappa in the comparison of models under the same approach, although there was an apparent similarity on model performance (Fig. 13). The smaller values achieved by the fuzzy Kappa, in relation to the juvenile and adult, could be related to the smaller on the radius of influence; the smaller radius is coherent with the spatial relations between fish, their field of vision and their competition, and it is logical that the mathematical effect of smoothing produces more similarities as the radius increases. The similarity between approaches was null, even providing negative values of the fuzzy Kappa statistic. The presence of negative values imply that the percentage of expected agreement by chance was larger than the real agreement; this effect is usual in 'natural' systems (personal communication Hagen, 2012) and was not considered an error. This results agreed with the preliminary observations (Fig 13), thus the unmodified Data-driven model determined fry as a pool dweller, in contrast with the Expert-knowledge models that relegated trout fry to shallower areas (Table 7 and Fig. 13).

The selection of the parameters in the calculation of the fuzzy Kappa should be adjusted in a trial and error procedure (personal communication, Hagen, 2012). However, some calculations were done in order to determine parameters as meaningful as possible, by selecting the average distance to the nearest trout of the proper size class, according to the density calculation. The density analysis considering each size class separately could be criticized because trout presents a relevant territorial behaviour (Chapman and Bjonn, 1969; Titus, 1990; Johnsson *et al.*, 2000; Brännäs et al., 2003), thus adults compete with juvenile and fry, and juvenile compete with fry.

Nevertheless, the results may suggest that the analysis was more generalist because adults and juveniles appeared concentrated over similar locations, although adult was more sparsely distributed. If the upper size class would had been used to calculate the trout density, the juvenile density would be concentrated it in a single point. Then, the correct assessment of that area would be probably too important. Finally, the analysis of the correlation between density and the suitability would be extremely conditioned by the proper assessment of that area. Regarding the fry, they appeared relatively far from juveniles and adults, but the density along the reach was higher, thus the mean radius was half the radius of adult and juvenile; as a consequence, no great differences could be expected if the analyses were carried out taking into account the older size classes.

5.6 Benefits of validation

The general overview of results corroborate the premise proposed by Guisan and Zimmermann (2000) about the need of the validation of habitat models with independent data to test its reliability and applicability. In the present case there were several issues to miss-select the proper models. Despite the existence of several studies that demonstrated difficulties in models' transferability (Fukuda, 2010), in general there should be no doubts about the validity of the data collected; rivers from the same region, with similar dimensions and an acceptable sample size, specially for adult and juvenile, but not for fry.

The flow assessed in the Cabriel River (Q=0.89 m^3/s) did not provide with the hydraulic conditions out of the surveyed extremes; the simulated flow produced a maximum average velocity of 0.53 m/s and a maximum depth of 1.4 m, both quite close to the maximum surveyed values in the Use datasets and clearly comprised in the Availability datasets. Therefore, both curves Category II ½ and Category III were within the range to develop fuzzy habitat suitability models; however, previous literature have reprobated Category III curves (Bovee, 1996). Then accordingly to that studies the wiser choice would had been the selection of the models based on the Category II ½ HSCs. Ignoring the option of the Category III HSCs and the Expert-knowledge models derived from. Additionally, the Data-driven approach has been widely and successfully applied (Mouton, 2008; Mouton *et al.*, 2009; Mouton *et al.*, 2011), but the cross-validation, especially in the fry case, has demonstrated to be insufficient to get the proper generalization capacity. Therefore the development of the Data-driven models unquestioning the obtained results could also derive in the selection of a worse or improper models. Altogether demonstrating the necessity about the development of proper validation strategies. Hence, in this case we recommend the Expert-knowledge fuzzy model based on Category III HSCs in the adult case, and the unmodified Data-driven fuzzy model in the juvenile case. The fry case should be based on an enlarged database but considering the actual models available, the best choice would be the Expert-knowledge fuzzy model based on Category III HSCs.

5.7 General comments

The Data-driven approach has demonstrated strongly dependent on the training database, but with a model validation its adjustment would be easily carried out. In addition this kind of models inform about the relations between the input variables and about how the variables jointly determine the rule consequents. However, the Expert-knowledge approach needs some assumptions or some data exploration which in most of the cases should be based on previous multivariate analysis which need its corresponding training database. In the Iberian Peninsula, the studies on fish habitat suitability are scarce due to the high percentage of endemisms and their limited distribution area (Ferreira et al., 2007). Most of that endemism are threatened (Smith and Darwall, 2006) thus leading to incomplete or 'imperfect' databases. Therefore in modelling habitat suitability for that uncommon species the field data collection could be unavoidable. In that sense, both methodologies could complement each other in a sort of feedback in order to improve model's reliability, in accordance with previous studies that emphasized the need of collaboration of both approaches (Mouton *et al.*, 2009).

The presented methodology allow modellers to generate Expert-knowledge fuzzy models derived from Habitat Suitability Curves (HSCs) on a systematic procedure, which is specifically recommended when no large experts panels are available. However, experts should be consulted if possible, and consulting companies should be discouraged of the generation and application of this kind of models by themselves if no experts are involved in their development. There is certain subjectivity on its development because the presented methodology includes some decisions and profanes could miss the correct choice; the author of this thesis was one of the persons involved in the data acquisition, which provides with some experience and knowledge on the field and fish behaviour.

In the present study the smoothing technique applied in the curve development was intended to be the minimum modification in order to get unimodal curves, but it incorporated certain subjectivity in the parameters selection. Further improvements will be focused on generating an objective and systematic smoothing procedure, allowing repeatability, which is considered basic on science. In addition, the expert judgment was also applied in the combinations of the partial suitability and in the determination of the controlling variables under some conditions, but its development over unknown species should be made with care, therefore experienced judgement should be involved in their development.

6 Conclusions

1. The Category II ½ Habitat Suitability Curves (HSCs) for the adult presented the highest suitability for low values of velocity and depth in comparison with some of the most relevant studies about adult brown trout. The Category III HSCs presented the optimum for faster flow and larger depth than most of the literature's curves. Instead, the substrate agreed most of them.

2. The Category II ½ HSCs for the juvenile did not presented a clear pattern in comparison with those from literature but the substrate which showed coincident with the optimum, over coarse substrate. The Category III HSCs coincided with most of their literature's counterparts.

3. The depth and velocity Category II ½ HSCs for the fry did not presented a clear pattern in comparison with those from literature. The major difference appeared in the substrate, the developed curve presented the optimum over silt whereas the literature showed it over coarse substrates. The depth and velocity Category III HSCs for the fry did not presented a clear pattern in comparison with those from literature. The major difference appeared also in the substrate curve.

4. The Expert-knowledge approach presented the capability to transform dichotomous input data into a wider range of outputs thus the output was categorised in three Fuzzy Sets.

5. The Expert-knowledge fuzzy models based on Category II ½ presented good sensitivity because assessed the trout locations with high suitability, although the specificity was low and in the adult and juvenile cases the suitability decreased in the most populated areas.

6. The Expert-knowledge fuzzy models based on Category III outperformed the based on Category II ½ thus presented good sensitivity and similar or better specificity. In addition all of them presented better correlation between trout density and suitability.

7. The Data-driven fuzzy models for adult and juvenile presented the better trade-off between sensitivity and specificity in addition to a positive correlation between density and suitability.

8. The Data-driven fuzzy models for fry presented the worse validation thus confirming the necessity of large-enough databases to properly apply the herein used Data-driven approach.

9. The Expert-knowledge approach showed consistent in the development of the Fuzzy Rules since no great differences with the Data-driven approach were observed. However the Expert-knowledge models based Category II ½ HSCs for adults and juveniles underrated the deeper areas.

10. The comparison of the model performance based on the fuzzy Kappa did not showed clear similarities between models neither intra-approaches nor inter-approaches.

11. The spatially explicit validation of the developed models has been demonstrated fundamental in the selection of the proper models for each of the considered size classes.

12. The better model for adult and fry were the Expert-knowledge models based on Category III HSCs whereas the best models for juvenile brown trout was the Data-driven fuzzy model.

7 Further research

Several future lines of research and possible improvements raised in accordance to the obtained results. The first one due to its simplicity was the necessity of enlargement of the brown trout fry database because it presented the smaller sample size and the most inconsistent results. This enlargement should consider also de presence of elder size stages, altogether improving the habitat suitability modelling. Additionally, this will offer a deeper insight into the discernment about the fry preferences for deep habitats reported by Raleigh (Raleigh, 1984) and the corresponding displacement due to the presence of elder life stages or rather than the opposite; its preference for shallower habitats. The second suggested improvement was the development of a systematic smoothing technique in the development of the HSCs aiming to provide unimodal curves but presenting the minimum possible alteration keeping as much as possible the original distribution. This could be carry out by means of an optimization algorithm; for instance the functional modification of the simplex algorithm (Rowan, 1990) implemented in R (R Development Core Team, 2012) by King (2008). A dichotomous output function that penalizes the multimodality can be easily developed in combining with the different functions for the mode calculation appeared in the R package 'modeest' (Poncet, 2012). The Data-driven approach showed affected by imperfections on the database (i.e. the presence uncovered rules or the impossibility to carry out real regression modelling). To improve that deficiencies alternatives to the hill-climbing algorithm could be tested. Recently a tool box have been developed in R (R Development Core Team, 2012); the 'frbs' package. This package host several alternatives (Riza *et al.*, 2013). Some of the most promising approaches are the HyFIS approach (Kim and Kasabov, 1999) and the Genetic Lateral Tuning of Linguistic Fuzzy

Systems (GLTLFS) (Alcalá *et al.*, 2007). The first one (HyFIS) belongs to the Fuzzy Neural Network (FNN) discipline which combines the human-like reasoning style of Fuzzy Inference Systems with the learning and connectionist structure of the Artificial Neural Networks (ANN) (Jang and Sun, 1995). HyFIS modifies the Membership Functions in to better predict the training data whereas the GLTLFS optimizes the Membership Functions through genetic algorithms but appearing coupled to a rule selection procedure thus automatically simplifying the Fuzzy Inference System. Finally, none of the aforementioned alternatives is stand-alone able to cope with low prevalence databases. Therefore, despite the possibility to carry out the subsampling procedure (Muñoz-Mas *et al.*, 2012), alternative techniques should be tested; for instance, the Probabilistic Neural Networks (PNN) (Specht, 1990). The PNN are a classificatory type of ANN which compare the assessed conditions with each datum included in the training database. To deal with differences on the intensity of the output, the weight of each category is inversely proportional to the number of training data in the corresponding category. Thus the classification to a given category depends on the values of the variables to determine the degree of membership to a given category but not on the amount of data from that category present in the training database.

8 Acknowledgements

This study was partially funded by the Spanish Ministry of Economy and Competitiveness with the project SCARCE (Consolider-Ingenio 2010 CSD2009-00065). Authors also thank the help of the Confederación Hidrográfica del Júcar (Gobierno de España) which provided environmental data, and to all the colleagues who collaborated in the field data collection.

9 References

Acreman, M., Dunbar, M.J., 2004. Defining environmental river flow requirements - A review. Hydrology and Earth System Sciences 8, 861-876.

Adriaenssens, V., De Baets, B., Goethals, P.L.M., De Pauw, N., 2004. Fuzzy rule-based models for decision support in ecosystem management. Science of the Total Environment 319, 1-12.

Ahmadi-Nedushan, B., St.-Hilaire, A., Bérubé, M., Ouarda, T.B.M.J., Robichaud, É., 2008. Instream flow determination using a multiple input fuzzy-based rule system: A case study. River Research and Applications 24, 279-292.

Alcalá, R., Alcalá-Fdez, J., Herrera, F., 2007. A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE Transactions on Fuzzy Systems 15, 616-635.

Alcaraz-Hernández, J.D., Martínez-Capel, F., Peredo, M., Hernández-Mascarell, A., 2011. Mesohabitat heterogeneity in four mediterranean streams of the Jucar river basin (Eastern Spain). Limnetica 30, 15.

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

Ayllón, D., Almodóvar, A., Nicola, G.G., Elvira, B., 2010. Ontogenetic and spatial variations in brown trout habitat selection. Ecology of Freshwater Fish 19, 420-432.

Bachman, R., 1984. Foraging Behavior of Free-Ranging Wild and Hatchery Brown Trout in a Stream. Transactions of the American Fisheries Society 113, 1-32.

Benaka, L.R.E., 1999. Fish habitat : Essential fish habitat and rehabilitation. Bethesda, Md. : American Fisheries Society 419.

Bovee, K.D., 1978. Probability of use criteria for the family Salmonidae. Information Paper #4 FWS/OBS-78/07. Fish and Wildlife Service, U.S. Department of the Interior., Washington, D.C., p. 53.

Bovee, K.D., 1982. A Guide to Stream Habitat Analysis Using the Instream Flow Incremental Methodology. Instream Flow Information, U.S. Fish and Wildlife Service, Fort Collins, Colorado., 248.

Bovee, K.D., 1986. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. Washington, DC: USDI Fish and Wildlife Service Instream Flow Information Paper #21 FWS/OBS-86/7., 235.

Bovee, K.D., 1996. The compleat IFIM : a coursebook for IF 250. U.S. Geological Survey, Biological Resources Division, Washington, D.C.

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

Bovee, K.D., Cochnauer., T., 1977. Development and evaluation of weighted criteria, probability of use curves for instream flow assessments: fisheries. Washington, D.C.: Fish and Wildlife Service, U.S. Department of the Interior Instream Flow Information Paper #3, FWS/OBS-77/63., 38.

Bovee, K.D., Lamb, B.L., Bartholow, J.M., Stalnaker, C.B., Taylor, J., Henriksen, J., 1998. Stream habitat analysis using the instream flow incremental methodology. Report USGS/BRD-1998-0004. U. S. Geological Survey, Biological Resources Division, Fort Collins, Colorado, p. 131.

Bovee, K.D., Zuboy, J.R., 1988. Proceedings of a workshop on the development and evaluation of habitat suitability criteria: A compilation of papers and discussions presented at Colorado State University, Fort Collins, Colorado, December 8-12, 1986. In: U.S. Department of the Interior, F.a.W.S. (Ed.), p. 407.

Brännäs, E., Jonsson, S., Lundqvist, H., 2003. Influence of food abundance on individual behaviour strategy and growth rate in juvenile brown trout (Salmo trutta). Canadian Journal of Zoology 81, 684-691.

Chapman, D.W., Bjonn, T.C., 1969. Distribution of Salmonids in Streams, with Special References to Food and Feeding. In: MacMillan, H.R. (Ed.), Symposium on Salmon and Trout in Streams. University of British Colombia, pp. 153-176.

Chen, Q., Mynett, A.E., 2003. Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. Ecological Modelling 162, 55-67.

Cock, M.J.W., 1978. The Assessment of Preference. Journal of Animal Ecology 47, 12.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20.

Elliott, J.M., 1973. The food of brown and rainbow trout (Salmo trutta and S. gairdneri) in relation to the abundance of drifting invertebrates in a mountain stream. Oecologia 12, 329-347.

Ferreira, T., Oliveira, J., Caiola, N., de Sostoa, A., Casals, F., Cortes, R., Economou, A., Zogaris, S., Garcia-Jalon, D., Ilhéu, M., Martinez-Capel, F., Pont, D., Rogers, C., Prenda, J., 2007. Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Management and Ecology 14, 473-481.

Fukuda, S., 2010. Assessing transferability of genetic algorithm-optimized fuzzy habitat preference models for Japanese Medaka (Oryzias Latipes). Mieres, pp. 57-62.

García, A., Jorde, K., Habit, E., Caamaño, D., Parra, O., 2011. Downstream environmental effects of dam operations: Changes in habitat quality for native fish species. River Research and Applications 27, 312-327.

Gatz Jr, A.J., Sale, M.J., Loar, J.M., 1987. Habitat shifts in rainbow trout: competitive influences of brown trout. Oecologia 74, 7-19.

Gibson, R.J., 1993. The Atlantic salmon in fresh water: spawning, rearing and production. Reviews in Fish Biology and Fisheries 3, 39-73.

Grossman, G.D., De Sostoa, A., 1994. Microhabitat use by fish in the lower Rio Matarrana, Spain, 1984-1987. Ecology of Freshwater Fish 3, 123-136.

Guay, J.C., Boisclair, D., Rioux, D., Leclerc, M., Lapointe, M., Legendre, P., 2000. Development and validation of numerical habitat models for juveniles of atlantic salmon (Salmo salar). Canadian Journal of Fisheries and Aquatic Sciences 57, 2065-2075.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147-186.

Hagen-Zanker, A., Straatman, B., Uljee, I., 2005. Further developments of a fuzzy set map comparison approach. International Journal of Geographical Information Science 19, 769-785.

Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information Science 17, 235-249.

Hagen, A., 2012. Recommendations to the proper selection of the decay function.

Hauer, C., Unfer, G., Graf, W., Leitner, P., Zeiringer, B., Habersack, H., 2012. Hydro-morphologically related variance in benthic drift and its importance for numerical habitat modelling. Hydrobiologia 683, 83-108.

Hayes, J.W., Jowett, I.G., 1994. Microhabitat models of large drift-feeding brown trout in three New Zealand rivers. North American Journal of Fisheries Management 14, 710-725.

Heggenes, J., 1990. Habitat utilization and preferences in juvenile Atlantic salmon (Salmo salar) in streams. Regulated Rivers - Research & Management 5, 341-354.

Heggenes, J., 1991. Comparisons of habitat availability and habitat use by an allopatric cohort of juvenile Atlantic salmon Salmo salar under conditions of low competition in a Norwegian stream. Holarctic Ecology 14, 51-62.

Heggenes, J., 1996. Habitat selection by Brown trout (Salmo trutta) and young atlantic salmon (S. SALAR) in streams: Static and dynamic hydraulic modelling. Regulated Rivers: Research and Management 12, 155-169.

Heggenes, J., Brabrand, Å., Saltveit, S., 1990. Comparison of Three Methods for Studies of Stream Habitat Use by Young Brown Trout and Atlantic Salmon. Transactions of the American Fisheries Society 119, 101-111.

Heggenes, J., Brabrand, Å., Saltveit, S.J., 1991. Microhabitat use by brown trout, Salmo trutta L. and Atlantic salmon, S. salar L., in a stream: a comparative study of underwater and river bank observations. Journal of Fish Biology 38, 259-266.

Jang, J.-S.R., Sun, C.-T., 1995. Neuro-fuzzy modeling and control. Proceedings of the IEEE 83, 378-406.

Johnson, H.D., 1980. The comparison of usage and availability measurement for evaluating resource preference. Ecology 61, 65-71.

Johnsson, J.I., Carlsson, M., Sundstrom, L.F., 2000. Habitat preference increases territorial defence in brown trout (Salmo trutta). Behavioral Ecology and Sociobiology 48, 373-377.

Jorde, K.S., M. Zoellner, F., 2000. Analysis of instream habitat quality-preference functions and fuzzy models., Stochastic Hydraulics 2000. Balkema, A.A., Rotterdam, pp. 241-248.

Jorde, K.S., Matthias. Peter, Armin. Zoellner, Frank., 2001. Fuzzy based Models for the Evaluation of Fish Habitat Quality and Instream Flow Assessment. Proceedings of the 2001 International Symposium on Environmental Hydraulics. Tempe, Arizona.

Jowett, I.G., Duncan, M.J., 2011. Effectiveness of 1D and 2D hydraulic models for instream habitat analysis in a braided river. Ecological Engineering.

Jutila, E., Ahvonen, A., Laamanen, M., 1999. Influence of environmental factors on the density and biomass of stocked brown trout, Salmo trutta L., parr in brooks affected by intensive forestry. Fisheries Management and Ecology 6, 195-205.

Kampichler, C., Barthel, J., Wieland, R., 2000. Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. Ecological Modelling 129, 87-99.

Karr, J.R., 1981. Assessment of Biotic Integrity Using Fish Communities. Fisheries 6, 21-27.

Kasabov, N.K., 1998. Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering. MIT Press, Cambridge, MA.

Kim, J., Kasabov, N., 1999. HyFIS: Adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. Neural Networks 12, 1301-1319.

King, A.A., 2008. subplex: Subplex optimization algorithm. R package version 1.1-3. http://CRAN.R-project.org/package=subplex.

Korman, J.P., C.J. Lekstrum, T., 1994. A Guide for the Selection of Standard Methods for Quantifying Sportfish Habitat Capability and Suitability in Streams and Lakes of B.C. Environment, Fisheries Branch, Research and Development Section, Vancouver, B.C., p. 87.

Lambert, T.R., Hanson, D.F., 1989. Development of habitat suitability criteria for trout in small streams. Regulated Rivers - Research & Management 3, 291-303.

Lamouroux, N., Capra, H., Pouilly, M., 1998. Predicting habitat suitability for lotic fish: linking statistical hydraulic models with multivariate habitat use models. Regulated Rivers: Research and Management 14, 1-11.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., Franklin, A.B., 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84, 2200-2207.

Magdaleno Mas, F., Martínez Romero, R., 2005. La lógica difusa en la modelización y mamejo de los ecosistemas acuáticos. Ingeniería Civil, 11.

Mäki-Petäys, A., Huusko, A., Erkinaro, J., Muotka, T., 2002. Transferability of habitat suitability criteria of juvenile Atlantic salmon (Salmo salar). Canadian Journal of Fisheries and Aquatic Sciences 59, 218-228.

Manel, S.C.W.H.C.W., H. Ormerod, S. J., 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. Journal of Applied Ecology 38, 921-931.

Martínez-Capel, F., Costa, R.M.S., Muñoz-Mas, R., Alcaraz-Hernández, J.D., Hernández-Mascarell, A.B., Patiño Rodríguez, J.H., Garófano-Gómez, V., 2009a. Factores de degradación de las poblaciones de loina (chondrostoma arrigonis) y el estado de su hábitat actual en la cuenca del río júcar (2006-2008). Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC) . Universitat Politècnica de Valè.ncia, Gandia (València). p. 169.

Martínez-Capel, F., García de Jalón, D., 1999. Desarrollo de Curvas de preferencia de microhábitat para Leuciscus pyrenaicus y Barbus bocagei por buceo en el río Jarama (Cuenca del Tajo). Limnetica 17, 4.

Martínez-Capel, F., García De Jalón, D., Werenitzky, D., Baeza, D., Rodilla-Alamá, M., 2009b. Microhabitat use by three endemic Iberian cyprinids in Mediterranean rivers (Tagus River Basin, Spain). Fisheries Management and Ecology 16, 52-60.

Martínez-Capel, F.P., M. Hernández-Mascarell, AB. Munné, T., 2008. Nose velocity calculation for spatial analysis of habitat and environmental flow assessments. 4th ECRR International Conference on River Restoration. ECRR European Centre for River Restoration, Venecia (Italia).

Michalewicz, Z., Fogel, D.B., 2000. How to Solve It: Modern Heuristics. Springer-Verlag, Berlin, Heilderberg, 467.

Mouton, A., Schneider, M., Goethals, P.L., De Pauw, N., 2007. Simulation of the physical habitat suitability of bullhead (Cottus gobio L.) in the Zwalm River. Communications in agricultural and applied biological sciences 72, 75-77.

Mouton, A.M., Alcaraz-Hernández, J.D., De Baets, B., Goethals, P.L.M., Martínez-Capel, F., 2011. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environmental Modelling and Software 26, 615-622.

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2009. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environmental Modelling and Software.

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2010a. Ecological relevance of performance criteria for species distribution models. Ecological Modelling 221, 1995-2002.

Mouton, A.M., De Baets, B., Peter, A., Holzer, G., Müller, R., Goethals, P.L.M.,, 2008. Entropy-based fuzzy set optimisation for reducing ecological model complexity. Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A., Guariso, G. (Eds.), iEMSs Fourth Biennial Meeting: “Summit on Environmental Modelling and Software”. International Modelling and Software Society/CD-ROM Publication, Barcelona, Spain.

Mouton, A.M., Dedecker, A.P., Lek, S., Goethals, P.L.M., 2010b. Selecting variables for habitat suitability of Asellus (Crustacea, Isopoda) by applying input variable contribution methods to artificial neural network models. Environmental Modeling and Assessment 15, 65-79.

Mouton, A.M., Schneider, M., Peter, A., Holzer, G., Müller, R., Goethals, P.L.M., De Pauw, N., 2008. Optimisation of a fuzzy physical habitat model for spawning European grayling (Thymallus thymallus L.) in the Aare river (Thun, Switzerland). Ecological Modelling 215, 122-132.

Moyle, P.B., Baltz, D.M., 1985. Microhabitat Use by an Assemblage of California Stream Fishes: Developing Criteria for Instream Flow Determinations. Transactions of the American Fisheries Society 114, 695-704.

Muñoz-Mas, R., Martínez-Capel, F., Schneider, M., Mouton, A.M., 2012. Assessment of brown trout habitat suitability in the Jucar River Basin (SPAIN): Comparison of data-driven approaches with fuzzy-logic models and univariate suitability curves. Science of the Total Environment 440, 123-131.

Nykänen, M., Huusko, A., 2004. Transferability of habitat preference criteria for larval European grayling (Thymallus thymallus). Canadian Journal of Fisheries and Aquatic Sciences 61, 185-192.

Orth, D.J., Maughan, O.E., 1982. Evaluation of the Incremental Methodology for Recommending Instream Flows for Fishes. Transactions of the American Fisheries Society 111, 413-445.

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

Poncet, P., 2012. modeest: Mode Estimation. R package version 2.1. http://CRAN.R-project.org/package=modeest.

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

Raleigh, R.F.Z., Laurence D.; Nelson, Patrick C., 1984. Habitat Suitability Index Models and Instream Flow Suitability Curves: Brown Trout. FWS/OBS - 82/10.124. U.S. Fish and Wildlife Service, p. 65.

Rincon, P.A., Lobon-Cervia, J., 1993. Microhabitat use by stream-resident brown trout: bioenergetic consequences. Transactions - American Fisheries Society 122, 575-587.

Riza, L.S., Bergmeir, C., Herrera, F., Benitez, J.M., 2013. frbs: Fuzzy Rule-based Systems for Classification and Regression Tasks. R package version 2.1-1. http://sci2s.ugr.es/dicits/software/FRBS.

Rosenfeld, J.S., Leiter, T., Lindner, G., Rothman, L., 2005. Food abundance and fish density alters habitat selection, growth, and habitat suitability curves for juvenile coho salmon (Oncorhynchus kisutch). Canadian Journal of Fisheries and Aquatic Sciences 62, 1691-1701.

Rowan, T.H., 1990. Functional Stability Analysis of Numerical Algorithms. University of Texas, Austin.

Rubec, P.J., J.C.W. Bexley, H. Norris, M.S. Coyne, M.E. Monaco, S.G. Smith, Ault J.S., 1999. Suitability modeling to delineate habitat essential to sustainable curves: white sucker. Fish and Wildlife Service. U.S. Department of the Interior. Washington, DC., 56.

Rust, S.W., Filgner, M.A., 1984. A modification of the kruskal-wallis statistic for the generalized behrens-fisher problem. Communications in Statistics - Theory and Methods 13, 2013-2027.

Saaty, T.L., 1980. The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill International Book Co., New York.

Savage, R.E., 1931. The relation between the feeding of the herring off the east coast of England and the plankton of the surrounding waters. Fishery Investigation, Ministry of Agriculture, Food and Fisheries, Series 2. 12, 88.

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

Service, U.S.F.a.W., 1981. Standards for the Development of Habitat Suitability Index Models. In: Service., U.S.F.a.W. (Ed.), Washington, DC.

Shannon, C.E., Weaver, W., 1963. Mathematical theory of communication. University of Illinois Press, Urbana, Illinois.

Smith, K.G., Darwall, W.R.T. (Eds.), 2006. The status and distribution of freshwater fish endemic to the mediterranean basin. IUCN -The World Conservation Union, Gland, Switzerland/Cambridge, UK.

Specht, D.F., 1990. Probabilistic neural networks. Neural Networks 3, 109-118.

Terrell, J.W., 1984. Proceedings of a workshop on fish habitat suitability index models. Fish and Wildlife Service, U.S. Department of the Interior Biological Report 85, 393.

Thomas, J.A., Bovee, K.D., 1993. Application and testing of a procedure to evaluate transferability of habitat suitability criteria. Regulated Rivers: Research & Management 8, 285-294.

Titus, R.G., 1990. Territorial behavior and its role in population regulation of young brown trout (Salmo trutta): new perspectives. Annales Zoologici Fennici 27, 119-130.

Vadas R.L, Jr., Orth, D.J., 2001. Formulation of habitat suitability models for stream fish guilds: Do the standard methods work? Transactions of the American Fisheries Society 130, 217-235.

Van Broekhoven, E., Adriaenssens, V., De Baets, B., Verdonschot, P.F.M., 2006. Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. Ecological Modelling 198, 71-84.

Vismara, R., Azzellino, A., Bosi, R., Crosa, G., Gentili, G., 2001. Habitat suitability curves for brown trout (Salmo trutta fario L.) in the River Adda, Northern Italy: Comparing univariate and multivariate approaches. River Research and Applications 17, 37-50.

Voos, K.A., 1981. Simulated Use of the Exponential Polynomial/maximum Likelihood Technique in Developing Suitability of Use Functions for Fish Habitat. Utah State University. Department of Civil and Environmental Engineering., Logan.

Wańkowski, J.W.J., Thorpe, J.E., 1979. Spatial distribution and feeding in atlantic salmon, Salmo salar L. juveniles. Journal of Fish Biology 14, 239-247.

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

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

Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338-353.