



Predictive models of fish microhabitat selection in multiple sites accounting for abundance overdispersion

Laura Plichard, Maxence Forcellini, Yann Le Coarer, Hervé Capra, Georges Carrel, René Ecochard, Nicolas Lamouroux

► To cite this version:

Laura Plichard, Maxence Forcellini, Yann Le Coarer, Hervé Capra, Georges Carrel, et al.. Predictive models of fish microhabitat selection in multiple sites accounting for abundance overdispersion. *River Research and Applications*, 2020, 36 (7), pp.1056-1075. 10.1002/rra.3631 . hal-02915577

HAL Id: hal-02915577

<https://hal.science/hal-02915577>

Submitted on 12 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PREDICTIVE MODELS OF FISH MICROHABITAT SELECTION IN MULTIPLE SITES ACCOUNTING FOR ABUNDANCE OVERDISPERSION.

3

4 Running head: PREDICTIVE MODELS OF FISH MICROHABITAT SELECTION IN MULTIPLE SITES

5

6 Authors: Plichard Laura^{1*}, Forcellini Maxence¹, Le Coarer Yann², Capra Hervé¹, Carrel Georges²,
7 Ecochard René³, Lamouroux Nicolas¹

8

9 ¹ Irstea, UR RiverLy, Villeurbanne, France

10 ² Irstea, UR RECOVER, Aix-en-Provence, France

11 ³ Université Lyon 1, Department of Health Biostatistics, CNRS 5558 - LBBE, Lyon, France

12 * Current address: Ecogea, 352 avenue Roger Tissandié 31600 Muret, France:

13 laura.plichard@protonmail.com

14

15 Forcellini Maxence : maxence.forcellini@irstea.fr

16 Le Coarer Yann : yann.lecoarer@irstea.fr

17 Capra Hervé : herve.capra@irstea.fr

18 Carrel Georges : georges.carrel@irstea.fr

19 Ecochard René : rene.ecochard@chu-lyon.fr

20 Lamouroux Nicolas : nicolas.lamouroux@irstea.fr

21

22 ACKNOWLEDGEMENTS

23 For financial support, thanks to Agence Française pour la Biodiversité, Compagnie Nationale du
24 Rhône, Agence de l'eau Rhône-Méditerranée-Corse, Electricité De France, Région Rhône-Alpes,
25 Région Provence-Alpes-Côte d'Azur. For fish datasets, thanks to the LTSER platform « ZABR »,
26 University of Marseille, Ecotec, and all contributors who helped for numerous sampling surveys. For
27 Bayesian statistic and simulation guidance, thanks to Dominique Lamonica. For parallel computing

advice, thanks to Gérard Henry, Dominique Lamonica, Julie Crabot and Éliane Schermer. This manuscript benefited from the meaningful comments from three anonymous reviewers.

30

Conflict of interest: No

32

ABSTRACT

Microhabitat selection models are frequently used in rivers to evaluate anthropogenic effects on aquatic organisms. Fish models are generally developed from few rivers, with debatable statistical treatments for coping with overdispersed abundance distributions. Analyses of data from multiple rivers are needed to test their transferability and increase their relevance for stakeholders. Using 3,528 microhabitats sampled in 9 French rivers during 129 surveys, we developed models for 35 specific-size classes of 22 fish species. We used mixed-effects GLMs (accounting for multiple surveys), involving B-spline transformations (accounting for non-linear responses) and assuming a Negative Binomial distribution (accounting for abundance overdispersion). We compared models of increasing complexity: no selection (M1), an "average" selection similar in all surveys (M2), two models with different selection across surveys (M3-4). Of 132 univariate cases (specific-size classes by habitat), 63% indicated selection for depth, 71% for velocity, 45% for substratum size and 13% for substratum heterogeneity. A total of 50 models were retained, involving 26/35 specific-size classes. Model fits indicated low explained deviance ($R^2_{MF} < 0.19$) and higher rank correlations ($\rho < 0.69$) between observed and modeled values. However, Bayesian posterior predictive checks validated these results since excellent fits would generate R^2_{MF} lower than 0.59 and ρ lower than 0.78. We found high transferability among rivers and dates, because (1) M2 was the most appropriate in 26/50 cases; (2) the R^2_{MF} and ρ values by M2 was respectively 72% and 75% of that explained by the complex M4, and (3) independent river cross-validations showed good transferability. Bivariate models for selected specific-size classes improved univariate model fits (ρ from 0.30 to 0.38). Overall, using a non-linear mixed-effect approach, our results confirmed the relevance of "average" models based on several rivers for developing helpful e-flow tools. Finally, our modeling approach opens opportunities for integrating additional effects as the spatial distribution of competitors.

56

Keyword: Fish microhabitat selection modeling, Fish preference, mixed-effect models, abundance overdispersion, Negative Binomial distribution, hydraulic habitat

59 INTRODUCTION

60 Habitat selection is the process by which an organism will choose the place to live (Odum, 1953;
61 Udvardy, 1959; Johnson, 1980). This choice will depend on several factors such as species or individual
62 traits (Wagner *et al.*, 2011; Fisher, Anholt & Volpe, 2011; Bunce *et al.*, 2013; Capra *et al.*, 2017). In order
63 to optimize their fitness (Fretwell & Lucas, 1969), organisms will select different types of habitat
64 according to their activities (e.g., foraging, reproduction period; Baker, 2006; Conallin *et al.*, 2014). To
65 describe this process, many habitat selection models are developed in terrestrial and in aquatic systems
66 (Manly *et al.* 2002). These models generally relate species abundance or occurrence to a variety of
67 habitat characteristics. They often involve habitat descriptions at the "local scale" of organisms, generally
68 referred to as microhabitats when the scale represents the immediate surroundings of organisms (Odum,
69 1953), or mesohabitats when the scale represents somewhat larger environments considered to be
70 functional units where individuals perform different activities (Kemp, Harper & Crosa, 1999; Fausch *et*
71 *al.*, 2002).

72 Habitat selection models are particularly needed in rivers, where the increase in water demand and
73 global warming are altering the local environment of freshwater organisms. In particular, the
74 multiplication of water withdrawal, weirs or dams (Lehner *et al.*, 2011) may affect ecosystem functioning
75 due to altered flow discharge, hydraulics, river morphology, habitat distribution and availability, and
76 dispersal possibilities (Poff, 1997; Vörösmarty *et al.*, 2010; Olden *et al.*, 2014). The development of
77 habitat selection models, in combination with hydraulic models and integrated into water management
78 software, has contributed to understand the variations in habitat suitability (often for fish and
79 macroinvertebrates) as a function of flow alteration (e.g., Tomsic *et al.*, 2007; Hayes, Hughes & Kelly,
80 2007; Conallin, Boegh & Jensen, 2010; Poff *et al.*, 2010; Lamouroux *et al.*, 2013; Garbe & Beeyers,
81 2017; Rosenfeld, 2017). Probably due to the strong variations in local hydraulics within rivers and their
82 strong dependence on discharge, many fish habitat selection models were developed at the microhabitat
83 scale (e.g., Lamouroux *et al.*, 1999; Mouton *et al.*, 2012; Booker & Graynoth, 2013), which represent the
84 immediate and daily habitat occupied by fish (Odum, 1953), or at the mesohabitat scale (e.g., Gosselin,
85 Maddock & Petts, 2012; Booker & Graynoth, 2013; Vezza *et al.*, 2014), which represent the functional
86 habitat for fish activities (e.g., pools and riffles, Kemp, Harper & Crosa, 1999). Thus, microhabitat and
87 mesohabitat models allow studying local processes involved in fish ecology, considering fish sensitivity
88 to local hydraulic variations. These models generally link species abundance (or other proxies of habitat
89 selection) to hydraulic variables such as water depth or current velocity, considered to be both more
90 direct and ecologically relevant descriptors of habitat conditions than flow discharge alone.

91 Numerous microhabitat selection models have been developed for fish, differing for example by their
92 localization (e.g., Labonne, Allouche & Gaudin, 2003; Nykanen & Huusko, 2003 in Europe, Costa *et al.*,
93 2013; Rosenfeld, 2017 in America; Fukuda, 2011; Shiroyama & Yoshimura, 2016 in Asia), the type of

94 rivers (e.g., Girard *et al.*, 2014 for tropical streams, Papadaki *et al.*, 2017 for Mediterranean streams) and
95 species involved (e.g., Dunbar, Alfredsen & Harby, 2012 for salmonids, Muñoz-Mas *et al.*, 2017 for redfin
96 barbels), or the methodological approach used (e.g., Radinger, Kail & Wolter, 2017 using expert
97 judgments; Jowett & Davey, 2007 using Generalized Additive Models, GAM). In many studies,
98 microhabitat selection models involved one or a few species and a limited number of surveys (dates ×
99 sites combinations). This often limits the possibility to test the transferability of models to independent
100 rivers of different catchments and to apply habitat models at the larger scale of “riverscapes” (Harby *et*
101 *al.* 2017). For that purpose, it is crucial to improve methodological developments of such models for
102 enlarging their application in river management.

103 The joint analysis of ecological data collected from multiple surveys is often useful to detect general
104 ecological patterns, and to account for the influence of various environmental parameters that can
105 influence patterns observed in each individual survey. For example, having numerous field surveys
106 increases the statistical power to detect changes after restoration projects (Vaudor *et al.*, 2015) and the
107 performance of species distribution models (Wisz *et al.*, 2008). For microhabitat selection models,
108 analyzing numerous surveys can increase the statistical power to detect similarities and differences in
109 microhabitat selection among surveys. Using numerous surveys is also required for non-abundant
110 species, for which a single survey is often not powerful enough for identifying robust abundance-habitat
111 relationships. However, a challenge when using data from several surveys is to develop modeling
112 frameworks that clearly separate abundance/occurrence changes within surveys (that correspond to
113 microhabitat selection) from differences between surveys that can be due to numerous other factors
114 such as temperature, water quality, hydrology or species biogeography.

115 Among the abundant literature on habitat selection, many statistical models have been used to study
116 microhabitat selection (Ahmadi-Nedushan *et al.*, 2006; Conallin, Boegh & Jensen, 2010). This includes
117 the simple comparison of microhabitat densities across habitat categories (e.g. habitat suitability curves
118 of Lamouroux *et al.*, 1999; Mouton *et al.*, 2012), generalized linear models (GLMs, e.g. Labonne,
119 Allouche & Gaudin, 2003; Jowett & Davey, 2007; Alcaraz-Hernandez *et al.*, 2016), fuzzy-models (e.g.,
120 Muñoz-Mas *et al.*, 2012) that compute a weighted average of different models, and more recent
121 machine-learning techniques such as random forests (e.g., Vezza *et al.*, 2014; Shiroyama & Yoshimura,
122 2016) or neural networks (e.g., Fukuda, 2011; Muñoz-Mas *et al.*, 2018) that are complex non-parametric
123 classification methods (Guisan & Zimmermann, 2000).

124 Due to their simplicity and flexibility, GLMs are powerful methods to assess fish habitat selection.
125 These models can be designed to cope for non-linear habitat selection relationships, for example by
126 using splines in generalized additive models (Pleydell & Chrétien, 2010; Girard *et al.*, 2014, Zuur,
127 Saveliev & Ieno, 2014). These smoothing regressions use polynomial regressions to characterize the
128 range of habitat selected without overfitting models. In addition, they can deal with the spatial and
129 temporal variations generated by multiple surveys within multiple rivers using mixed-effect formulations

130 (Goldstein, 2003; Bates *et al.*, 2015). They also can account for different hypothesis concerning the
131 statistical distribution of abundance counts, such as the Poisson or Negative Binomial distributions.
132 Because GLMs are parametric models, their parameters can be extracted for predictive purposes
133 (Ahmadi-Nedushan *et al.*, 2006) and easily transferred to stakeholders as “preference curves” (e.g.,).
134 Combinations of multilevel GLMs and smoothing B-splines have already been used with longitudinal
135 data (i.e. multiple measurements on the same subject) in medicine, for modeling human growth (Pan &
136 Goldstein, 1998; Grajeda *et al.*, 2016), but in our knowledge they have not been used in ecology for
137 modeling the relationship between species abundance and habitats. By combining a regression method
138 with a smoothing approach, this modeling method could help developing a powerful method for studying
139 complex microhabitat selection processes.

140 However, the overdispersion of fish abundance data (i.e. variance \gg mean) complicates
141 microhabitat selection modeling. This overdispersion is characterized by a high number of samples
142 without organisms and a few samples containing the majority of the abundance. This overdispersion can
143 be accounted for by different methods. A first possibility is to transform the data (e.g., by log-
144 transformation; O'Hara & Kotze, 2010) in order to limit the impact of high variance on the models, but
145 transformations generate difficulties such as an underestimation of mean values (O'Hara & Kotze, 2010;
146 Warton, Wright & Wang, 2012; Ives, 2015). Another method is to use distribution assumptions that
147 involve an overdispersion parameter by definition, such as the overdispersed Poisson distribution, the
148 Negative Binomial distribution or zero-inflated distributions, which consider simultaneously an
149 abundance distribution and a presence-absence distribution (Warton, 2005; Potts & Elith, 2006; Vaudor
150 *et al.*, 2015). Of these distributions, the Negative Binomial is particularly well adapted to freshwater
151 abundance data (Vaudor, Lamouroux & Olivier, 2011) and might be the most relevant for modeling fish
152 microhabitat selection.

153 Given these statistical limits of current approaches, the aim of this paper is to present and apply a
154 more suitable methodology for developing fish microhabitat selection models. Our approach is based on
155 the use of mixed-effect GLMs (to account for data collected in multiple surveys) involving B-spline
156 transformations of habitat variables (to account for non-linear responses) and Negative Binomial
157 assumptions (to account for overdispersed abundance data). We applied our approach to a unique data
158 set involving 3,528 microhabitats electrofished in a total of 129 surveys (date \times reach combinations) in 9
159 French rivers. A total of 22 species \times four size classes were considered. We first developed univariate
160 models relating fish abundance to microhabitat water depth, current velocity, substratum grain size and
161 substratum heterogeneity (i.e. grain size diversity). Models of increasing complexity were compared to
162 appreciate the degree of model transferability across surveys, and models were built for microhabitat
163 subsets with or without hydraulic refuges. Their performance was assessed using several metrics,
164 posterior predictive checks (Bayesian approach) and cross validations. Then, we developed bivariate

models for a selection of species to appreciate their added values when compared with univariate models.

METHODS

STUDY REACHES

A total of 3,528 microhabitats were sampled between 1989 and 2014 during 129 surveys in 9 southern French rivers (Fig. 1, Table 1, Appendix 1). A first dataset was (n= 1709 microhabitats; rivers Ain, Ardèche, Drome, Garonne, Loire and Rhône) was already used by Lamouroux *et al.* (1999) to develop habitat selection models, with a different modeling approach and a debatable log-transformation of abundance. A second dataset (n= 1819 microhabitats) involved southern rivers. These two datasets were pooled to develop fish microhabitat selection models.

FISH SAMPLING

In both datasets, fish were electrofished at low flow rates with adjusted current (5 kW, 180-1000 V, 1-4 A, direct current) in microhabitats of varying surface areas (between 4 and 90 m² for 95% of microhabitats). Each microhabitat was selected in a given habitat type (e.g. a pool or a riffle) and electrofished using an open-sampling technique (Vadas & Orth, 1993; Lamouroux *et al.*, 1999). For each survey, the total number of microhabitats represented the distribution of habitat types observed in the sampling reach on the sampling date. Fish lengths were recorded and classified into four classes cl1 (\leq 80 mm), cl2 (80 – 180 mm), cl3 (180 - 300 mm) and cl4 ($>$ 300 mm).

To model fish microhabitat selection, for each specific size class (i.e. a size class associated to one species), surveys including less than six individuals observed in less than three microhabitats were removed from datasets. In addition, species observed in less than three surveys were not modelled. Some classes were pooled to develop and improve the models when individual classes did not contain enough individuals. For example, a pooled class 1-3 will be noted cl123 hereafter.

MICROHABITAT CHARACTERISTICS

Mean water depth and mean water column current velocity were calculated by the mean of several verticals measurements (average = 7), randomly sampled and georeferenced within each microhabitat, defining the microhabitat and weighted by their representative area for depth and representative volume for velocity. Point water column velocities were measured from three measurements at distances above the bed of 20%, 40% and 80% of the water depth, using a current meter. These calculations were performed using the HydroSignature software (Le Coarer, 2007).

197

198 Substratum grain size characterisation differed between the two datasets. In the first dataset, one or two
199 dominant substratum grain sizes (covering most of the $\sim 1\text{m}^2$ area around the point) and the substratum
200 with the maximum grain sizes were estimated at several points, as for depth and velocity, and assigned
201 to one of 12 categories using a modified version of the Wentworth logarithmic scale (Wentworth, 1922).
202 In the second dataset, the substratum grain size was measured at 10 points using the roughness height
203 of substratum, defined as the relative height of particles relative to the bed (Gordon, McMahon &
204 Finlayson, 1992). To merge datasets, we translated all substratum characterisations into frequencies of
205 seven size classes within the microhabitat: silt, sand, gravel, pebble, cobble, boulder and block. Then,
206 we computed the dominant substratum class and the number of distinct substratum classes. Finally, for
207 each species, between 29% and 50 % of the data did not include substratum descriptions and were
208 removed from dataset for developing models involving substratum (Table 2).

209 The presence of hydraulic refuges (grouping mineral refuges, vegetated refuges and bank refuges) was
210 recorded. We separated the full dataset into two datasets, with or without refuges, to analyse the
211 influence of refuges on microhabitat use. Most analyses were made on the dataset without refuges, and
212 the data set with refuge was used to appreciate their influence.

213

214 MODELING

215 To compare predictive microhabitat selection models by specific size classes (and refuge types) we
216 defined four mixed-effect models of increasing complexity (M1-M4, eqn.1 - eqn.4). In short, M1 was a
217 model without microhabitat selection, M2 a model where selection was similar in all surveys and M3-4
218 models where selection could vary across surveys. All models linked the abundance of a specific size
219 class to one of our four habitat variables, and accounted for difference between surveys by forcing a
220 random effect on the model intercept. Models M3-4 also had random effects at the survey level
221 associated with microhabitat characteristics, allowing variable microhabitat selection across surveys.
222 Each model could be represented by its conditional expression, which corresponded to the full model fit
223 (fixed and random effects) and thus could vary across surveys (e.g., Fig. 2), and its marginal component,
224 which corresponded to its fixed effects only and had a similar form across surveys (e.g., Fig. 3).

225 Fish overdispersion was accounted for by assuming that abundance values followed a Negative
226 Binomial distribution $\text{NB}(\mu, \theta)$, with different parameters of dispersion θ across species, size class and
227 models, but a constant θ across surveys. Vaudor *et al.* (2015) justified the choice of using a constant
228 parameter across surveys by the reduced temporal variation of θ for a given species. To check the
229 relevance of our NB assumption, we estimated for each model the overdispersion coefficient defined as
230 the sum of squared Pearson residuals divided by the difference between the sample size and the
231 number of parameters (Zuur, Saveliev & Ieno, 2014). The values, close to 1, ranged between 0.4 and

232 2.4 and indicated a good adjustment of the NB to our abundance distributions (Zuur, Saveliev & Ieno,
233 2014).

234 Because microhabitat selection is typically a non-linear process (e.g., Labonne, Allouche & Gaudin,
235 2003; Girard *et al.*, 2014; Alcaraz-Hernandez *et al.*, 2016), we introduced B-splines to transform the
236 microhabitat variables, which decomposed these variables into piecewise cubic regressions with fixed
237 knots (Pan & Goldstein, 1998; Grajeda *et al.*, 2016). For each model, we previously selected the
238 appropriate number of knots between models with a single knot (fixed at 50 % of the distribution of the
239 microhabitat variable, 2 degrees of freedom) or models with two knots (fixed at 33 % and 67 % of the
240 distribution of the microhabitat variable, 3 degrees of freedom). The criterion used for this selection was
241 a modified Akaike Information Criterion for overdispersed count data, QAIC (Burnham & Anderson,
242 2002; Kim *et al.*, 2014), which is a correction of the maximum-likelihood estimations by the
243 overdispersed coefficient of the global model and is calculated as:

$$244 \quad QAIC = -2 \frac{\log(Likelihood)}{\hat{c}_{M4}} + 2K$$

245 where K is the number of parameters of the model, and \hat{c}_{M4} the overdispersion coefficient estimated
246 using our more detailed model M4.

247 Finally, to deal with differences in surface areas among microhabitats, which obviously could influence
248 abundance values, we introduced an offset (c_{area}) which was not a parameter of the models and
249 corresponded to the total area sampled by microhabitat.

250 The four models were built following:

$$251 \quad y_{ij} \sim NB(\mu_{ij}, \theta)$$

252 where y_{ij} is the abundance of a specific size class of microhabitat i in survey j and μ_{ij} its expected mean
253 value. Relations between μ_{ij} and microhabitat variables varied across models:

254

255 **M1: No microhabitat selection**

$$256 \quad \log(\hat{\mu}_{ij}) = \beta_0 + u_{0j} + c_{area} \quad \text{eqn.1}$$

257 where β_0 represents the fixed component of the intercept and $u_{0j} \sim N(0, \sigma_0^2)$ its random component.
258 Values of u_{0j} were assumed normally distributed with a standard deviation of σ_0 .

259 M1 is a model where microhabitat variables have no influence on abundance.

260

261 **M2: “Average” microhabitat selection**

262 $\log(\hat{\mu}_{ij}) = \beta_0 + u_{0j} + \beta_{x_{t_n}} * f(x_{t_n}, n_{ij}) + c_{area}$ eqn.2

263 where $\beta_{x_{t_n}}$ are the fixed coefficients for each cubic regression spline $f(x_{t_n}, n_{ij})$ of the microhabitat
264 variable x with n knots located at t_n positions, and $u_{0j} \sim N(0, \sigma_0^2)$.

265 M2 is a model where microhabitat variables have a similar influence on abundance across surveys, i.e. it
266 assumes that microhabitat selection is transferable across sites and dates.

267

268 **M3: Partially random microhabitat selection**

269 $\log(\hat{\mu}_{ij}) = \beta_0 + u_{0j} + \beta_{x_{t_n}} * f(x_{t_n}, n_{ij}) + u_{1j} x_{ij} + c_{area}$ eqn.3

270 where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(0, \Omega)$ with a variance covariance matrix $\Omega = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$.

271 M3 is a model where microhabitat variables have different influences on abundance across surveys.

272

273 **M4: Fully random microhabitat selection**

274 $\log(\hat{\mu}_{ij}) = \beta_0 + u_{0j} + (\beta_{x_{t_n}} + u_{1x_{t_n}}) * f(x_{t_n}, n_{ij}) + c_{area}$ eqn.4

275 where $\begin{pmatrix} u_{0j} \\ u_{1x_{t_n}} \end{pmatrix} \sim N(0, \Omega)$ and $\Omega = \begin{pmatrix} \sigma_0^2 & \cdots & \sigma_{0x_{t_n}} \\ \vdots & \ddots & \vdots \\ \sigma_{0x_{t_n}} & \cdots & \sigma_{x_{t_n}}^2 \end{pmatrix}$.

276

277 M4 is a model where microhabitat variables have different influences on abundance across surveys as in
278 M3, but is more flexible than M3 because all coefficients of cubic regression splines may vary across
279 surveys.

280 In our study, M1 is considered as a null random model and thus a reference for appreciating the added
281 value of microhabitat selection models M2-M4. By contrast, M4 is the most complicated and
282 parametrized model and represents the best fitted microhabitat selection model.

283

284 **MODEL EVALUATION**

285 *Model selection*

286 To compare models fit and parsimony, we calculated a ΔQAIC , defined as the difference in QAIC
287 between the model with the lowest QAIC and the QAIC of the three other models. Following Burnham &

288 Anderson (2002), models presenting a $\Delta QAIC \geq 10$ were identified as failing to explain some substantial
289 variation in the data.

290 *Model fits: explained deviance (R^2MF) and Spearman rho*

291 Model fits were characterized using a McFadden's R^2 (R^2MF ; McFadden, 1974) or explained deviance,
292 which is the ratio between the explained deviance and the null deviance and is the equivalent for
293 maximum-likelihood of a R^2 in linear regression (Guisan & Zimmermann, 2000). Because R^2MF are
294 based on null deviance, they can be compared between each model. A model with a higher value of
295 R^2MF indicates a model with a higher explained deviance.

296
$$R^2MF = 1 - \frac{\log(Likelihood_{Model})}{\log(Likelihood_{Nullmodel})}$$

297 To appreciate how the fully random model (M4) fitted the data compared to the model without
298 microhabitat selection (M1), we calculated a ΔR^2MF which compared the difference between the R^2MF of
299 M4 and the R^2MF of M1.

300

301 To appreciate how the “average” model M2 (considered as an average selection model shared by all
302 rivers) explained microhabitat selection compared to the most flexible model (M4), we calculated a
303 relative R^2MF ($RRMF$), which compared the differences between the R^2MF of the model M2 and the
304 R^2MF of the simplest model M1, to the difference between the R^2MF of the most parametrized model
305 (M4) and the least parametrized model (M1).

306
$$RRMF_{M2} = \frac{R^2MF_{M2} - R^2MF_{M1}}{R^2MF_{M4} - R^2MF_{M1}}$$

307 We also use Spearman rho (i.e. rank correlation; Spearman, 1904), hereafter cited as Spearman rho, as
308 an alternative statistic to R^2MF , in order to complement our appreciation of model fit (Guisan &
309 Zimmermann, 2000; Potts & Elith, 2006). Adding this statistic was important in our case of overdispersed
310 abundance values, which cause low R^2MF values. Spearman rho is the correlation between the rank of
311 observed values and the rank of conditional fitted values (i.e. model adjusted by survey). A value close
312 to 1 suggests a positive correlation between the rank of observed values and the rank of conditional
313 fitted values, consequently a good rank correlation between the observed values and the predictions of
314 the observed values by the conditional model. By contrast, negative value of Spearman rho correlation
315 indicates negative correlation between the rank of observed values and the rank of fitted values,
316 consequently a poor rank correlation between the observed values and the predictions of the observed
317 values by the model. As for R^2MF , we calculated a Δrho which compared the difference between the
318 Spearman rho of M4 and the Spearman rho of M1.

319

320 Again, to appreciate the part of microhabitat selection explained by the "average" fixed model M2, we
321 calculated a relative Spearman rho, which compared the differences between the Spearman rho of
322 model M2 and the Spearman rho of the simplest model M1, to the difference between the Spearman rho
323 of the most parametrized model (M4) and the least parametrized model (M1).

324
$$Rrho_{M_2} = \frac{rho_{M_2} - rho_{M_1}}{rho_{M_4} - rho_{M_1}}$$

325 *Posterior predictive check simulations*

326 We used posterior predictive check simulations (Chambert, Rotella & Higgs, 2014), widely used in
327 Bayesian statistics, to assess the performance of models at predicting the data. The objective of this
328 technique was to appreciate what quality of model fit (i.e. R²MF and Spearman rho) could be expectable
329 with overdispersed abundance data such as ours. Specifically, simulations consisted in 1) assuming that
330 the fitted model corresponded to the true response of abundance to habitat, 2) generating fictive,
331 simulated observed values (n = 1500 fictive abundance datasets) taking into account the overdispersion
332 of the taxa, 3) estimating the R²MF and Spearman rho of models fitted on these fictive data, and 4)
333 comparing our initial model fit with the fictive ones.

334 The choice of n = 1500 fictive datasets was made after trials with n values between 20 and 2000,
335 indicating stable results for all models when n = 1500. To keep a similar structure between the observed
336 dataset and the simulated datasets, each simulated dataset contained the same number of surveys as
337 the observed dataset. In a simulated survey, we randomly picked the number of microhabitats within the
338 range of microhabitats sampled by survey. For example, *Telestes soufia* class 2 was present in 60
339 surveys which contained between 7 and 96 microhabitats, consequently the simulated surveys randomly
340 contained between 7 and 96 microhabitats. Then we randomly picked a value of the microhabitat
341 characteristic for each sample. Given this microhabitat characteristic, we finally randomly picked the
342 sampling area within the range of area observed for this microhabitat value. For example, the value of
343 0.1 m of water depth was observed in sampling area between 2 and 94 m² with a median at 21 m².
344 Finally, the abundance \hat{y}_{ij} was simulated following $NB(\hat{\mu}, \hat{\theta})$ using the 95 % confidence interval
345 estimations for fixed effect and the variance estimated for the random effect from models. Because M4 is
346 assumed to represent all the variability due to the hydraulic variables measured, we used in the
347 simulations the $\hat{\theta}$ estimated from M4 for all simulations (M1-M4).

348

349 *Leave_one_river_out cross-validations*

350 To evaluate the transferability of M2 across rivers, i.e. its ability to predict microhabitat selection on
351 external data, we computed leave_one_river_out cross-validations after removing each river in turn.
352 Specifically, for each specific class, we computed new models after excluding one river (i.e. training data
353 set) and calculated the Spearman rho correlations between the observations and the predictions of each
354 survey from this river (i.e. validation data set). The specific size classes included in these cross-
355 validations were those showing a minimum magnitude of microhabitat selection (subjectively chosen as
356 $\Delta R^2_{MF} > 0.01$ or $\Delta \rho > 0.09$ between M4 and M1) and well described in the data (subjectively chosen as
357 abundance ≥ 20 individuals per survey, occurrence ≥ 10 per survey).

358

359 *Comparison with the dataset with refuges and the results of Lamouroux et al. 1999*

360 The comparison of models for datasets with and without refuges was done for the selection of specific
361 size classes showing a minimum magnitude of microhabitat selection (see just above).

362 Then, we compared our selection models with preference models showed in Lamouroux *et al.* (1999),
363 derived with different methods (calculations of average log-densities in habitat classes in Lamouroux *et al.*
364 1999). For this purpose we represented the predicted abundance in our study (according to the
365 marginal model M2, which corresponded to an average model across survey comparable to those
366 developed in Lamouroux *et al.* 1999) as a function of the preferred microhabitat class identified in
367 Lamouroux *et al.* (1999). This comparison was made only for specific size classes showing a minimum
368 magnitude of microhabitat selection.

369

370 *Multivariate models*

371 To evaluate whether multivariate models increased the performance of univariate models, we developed
372 multivariate models that include an additive selection for water depth and current velocity. Multivariate
373 models were built for the "average" model M2 only and for specific size classes showing a minimum
374 magnitude of microhabitat selection. We compared the R^2_{MF} and rho of multivariate models with values
375 obtained by each univariate model (i.e. water depth and current velocity).

376 All modeling was performed with R software (version 3.4.1, R Core Team, 2017) using the function
377 `{glmer}` from lme4 package (Bates *et al.*, 2015) for mixed effects models, mgcv package (Wood, 2006)
378 for Negative Binomial distributions.

379 RESULTS

380 *Species characteristics*

A total of 87,177 individuals of 22 species, mainly cyprinids (Table 1), were considered in the dataset. They belonged to 37 specific size classes. The two most abundant species were *Barbatula barbatula* cl12 (N = 14,661 individuals) and *Phoxinus phoxinus* cl12 (N = 11,721), and the two least abundant species were *Zingel asper* cl12 (N = 29) and *Rhodeus amarus* cl12 (N = 66). The two species the most occurring were *Barbatula barbatula* cl12 (N = 1,124 presences in N = 2,359 sampled microhabitats) and *Phoxinus phoxinus* cl12 (N = 844 /1,854), and the two least occurring species were *Zingel asper* cl12 (N = 18 /196) and *Chondrostoma nasus* cl34 (N = 67 / 465). Around two third of specific size classes were sampled in more than 10 surveys.

Concerning the refuges, a quarter of specific size classes occurred more frequently in microhabitats containing refuges than others. Although refuge descriptions are not detailed here, the majority of refuges were vegetated refuges.

392

393 MODEL EVALUATION

394 *Model selection*

Models were constructed using the dataset without refuges for a total of 132 cases (size class x variable combinations) that involved 35 specific size classes for water depth and current velocity, and 31 size classes for substratum variables. Following model selection according to $\Delta Q A I C$ criteria, there was a significant microhabitat selection for 65/132 cases: 22/35 for water depth, 25/35 for current velocity, 14/31 for dominant substratum grain size, and 4/31 for substratum heterogeneity. In total, 18/22 species and 29/35 size classes significantly selected at least one habitat characteristic. In the remaining cases, model M1 was selected, suggesting no microhabitat selection.

Among the 65 significant selection models, we retained hereafter only the 50 cases (Table 3) with a minimal magnitude of habitat selection, subjectively defined as $\Delta R^2 M F$ between M4 and M1 strictly superior to 0.01 or a $\Delta \rho$ strictly superior to 0.09 (models listed in Table 3). These 50 cases involved 19/22 species and 26/35 specific size classes. For these models, M2 was selected in 26/50 cases (11 for water depth, 10 for current velocity, 4 for dominant substratum grain size, 1 for number of substratum classes), M3 in 21/50 cases (9 for water depth, 8 for current velocity, 3 for dominant substratum grain size, 1 for number of substratum classes) and M4 in 3/50 cases only (all for current velocity).

409

410 *Model fits: explained deviance ($R^2 M F$) and Spearman rho*

The values of $R^2 M F$ obtained were generally low (see Table 3 for a synthesis of model fits): between 0 and 0.10 for M1, from 0.02 to 0.16 for the “average” model M2, from 0.02 to 0.19 for the detailed M4. Because the level of complexity of M3 is between the level of complexity of M2 and M4, values for M3 were not shown in Table 3. When analyzing how the fully random model (M4) fitted the data compared to

the model without selection (M1), we observed differences in R^2MF from 0.02 to 0.17 (between 0.02 and 0.09 for water depth, 0.02 and 0.16 for current velocity, 0.02 and 0.17 for dominant substratum grain size, 0.02 and 0.08 for number of substratum classes).

When analyzing how the “average” model (M2) fitted the data compared to the fully random model (M4) using the relative R^2MF , we observed values from 0.15 to 1.03 with a median of 0.72. In other words, the “average” model M2 explained a median share of 72% of the deviance explained by the detailed model M4.

Conditional Spearman rho (Table 3) confirmed these global results but rho values were much higher than R^2MF values. Spearman rho obtained ranged from -0.25 to 0.38 for M1, from 0.04 to 0.67 for M2, and from 0.07 to 0.69 for M4. When analyzing how the fully random model (M4) fitted the data compared to the model without selection (M1), we observed differences in conditional Spearman rho from 0.10 to 0.62 (between 0.10 and 0.41 for water depth, 0.10 and 0.62 for current velocity, 0.11 and 0.51 for dominant substratum grain size, 0.10 and 0.41 for number of substratum classes).

When analyzing how the “average” model (M2) fitted the data compared to the fully random model (M4) using the relative conditional Spearman rho, we observed values from -0.24 to 1.01 with a median of 0.75. In other words, the “average” model M2 had rho values that represented a median share of 75% of the rho values of model M4, consistently with the results obtained with deviance statistics.

Posterior predictive check simulations

The posterior predictive check simulations (Table 3) suggested that, considering the overdispersion of microhabitat abundance data, the highest R^2MF and Spearman rho values (95% percentile of our fictive simulations) that could be expected when fitting our models ranged between 0 and 0.59 (for R^2MF) and 0.28 and 0.78 (for Spearman rho). Our model fits (up to 0.19 for R^2MF and up to 0.69 for Spearman rho) therefore indicates satisfactory fits for overdispersed data.

Univariate average microhabitat selection

Fig. 2 and Fig. 3 provide two examples of the univariate models described in Table 3. Fig. 2 shows how the different models M1-M4 fitted the observed variations in observed abundance of small *Barbatula barbatula* cl12 as a function of water depth and *Telestes soufia* cl1 as a function of current velocity. The *Barbatula barbatula* cl12 model has a strong magnitude of habitat selection, with a $\Delta\rho$ of 0.32, and the gain in Spearman rho with model M2 (Rrho) equals 93% of the gain in rho with M4 (Table 3). The “average” selection model is therefore very relevant for this size class. The *Telestes soufia* cl1 is an example with lower magnitude of habitat selection ($\Delta\rho$ = 0.26 and Rrho = 73%, current velocity,

Table 3). For these taxa, the “average” model is also relevant but variations in selection across surveys are stronger (deviations between the red-line M2 and black line M4). Both graphs enable to appreciate the high degree of dispersion in these microhabitat data, with many points without fish and others with >100 individuals. Fig. 3 shows the marginal component of M2 for these two taxa, shared by all surveys, that represents the average shape of habitat selection model for the whole dataset.

Fig. 4 summarizes the “average habitat selection” of habitat variables for all models in Table 3. This “average habitat selection” corresponds to the average value of the marginal component of M2 (as shown in the examples of Fig. 3) over the range of habitat characteristics observed in our dataset ([0, 3] m for water depth; [0, 1.94] m.s⁻¹ for water column current velocity). It corresponds to the average habitat value used by the size class if all habitat characteristics were equally available over these ranges. Fig. 4 indicates a wide diversity of habitat selection among specific size classes, with average selected depth ranging between ~0.2 and ~1.2 m, average selected velocities between ~0.05 and ~1 m.s⁻¹ and substratum sizes between sand and cobbles. For example, Fig. 4 indicates that small *Barbatula barbatula* cl12 select shallow habitats around ~0.2 m whereas *Telestes soufia* cl1 select intermediate velocities ~0.3 m.s⁻¹, consistently with Figs. 2&3. Other species such as *Barbus barbus* cl34 select deep and fast-flowing habitats, and large fish tend to select deeper habitats than small ones. Among the fewer significant selection of substrate characteristics, small *Perca fluviatilis* cl12 selects a reduced substratum heterogeneity (as opposed to *Squalius cephalus* cl2), preferably gravels.

Leave_one_river_out cross-validation

Cross-validations of the models of Table 3 indicated the majority of models M2 presented positive correlation between the observations and the predictions of the validating data set (Fig. 5). Cross-validation rho values were of the same order of values than the differences in Spearman rho between M2 and M1 models during the fit on all rivers. Therefore, they reflected the potential of model M2 to predict abundance ranks in surveys of independent rivers. This was particularly true for example *Barbatula barbatula* cl12 (water depth, described in Fig. 3), *Telestes soufia* cl1 and *Barbus barbus* cl2 (current velocity, described in Fig. 3) but less for *Alburnus alburnus* cl1 (water depth, Fig. 3) whose cross-validations were less convincing.

Comparison with the dataset with refuges and the results of Lamouroux et al. 1999

After fitting selection models for the dataset with refuges (as was done for the dataset without refuges), a total of 24 significant models could be compared between the two datasets: 6 for water depth, 17 for current velocity, 1 for dominant substratum grain size and none for substrate heterogeneity. For all habitat variables, the majority of specific size classes presented similar average selection for models build on data with and without refuges (Fig. 6). A few deviations were observed for *Barbus barbus* cl2

484 and cl34, and *Alburnoides bipunctatus* cl23 that used lower velocities in the presence of refuges, and for
485 *Squalius cephalus* cl34 that selected a smaller substratum (gravel vs. pebble) in presence of refuges.

486 When comparing microhabitat selection (models of Table 3) with those available in Lamouroux et al
487 1999 (Fig. 7), we observed consistent results for water depth and current velocity but not for dominant
488 substratum grain size.

489

490 *Multivariate models*

491 Multivariate models were developed for the 16 specific size classes that had significant selection (with
492 minimal magnitude of effect) for both water depth and current velocity (Table 3). The multivariate model
493 approach increased the performance of the majority of univariate models M2. This result was observed
494 with both metrics: median R^2_{MF} increased from 0.04 for water depth and 0.05 for current velocity to 0.06
495 and median ρ increased from 0.30 for water depth and 0.33 for current velocity to 0.38 (Fig. 8).

496 DISCUSSION

497 Our results confirmed the generality of microhabitat selection because 29/35 specific size classes
498 showed significant microhabitat selection for at least one hydraulic variable, with a minimal magnitude.
499 Moreover, our results indicated the relevance of average models across surveys because (1) the
500 average model (M2) was selected by Akaike criteria in 26/50 cases, and (2) across all cases, the
501 average model explained more than 70 % of the variability explained by a fully random microhabitat
502 selection model (i.e. random intercept and slope by survey, model M4). However, if this fully random
503 microhabitat selection model explained a small proportion of variance, then the average model would not
504 be relevant either. Our results were also consistent with previous studies on fish habitat selection (e.g.,
505 Riffart *et al.*, 2009), for example small *Barbatula barbatula* selected shallow microhabitats and small
506 *Barbus barbus* faster-flowing microhabitat.

507 In contrast to our expectations, no difference in microhabitat selection was found for both
508 conditions with and without refuge. Although refuges are crucial for fish ecology (e.g., Magoulick and
509 Kobza 2003), these results could suggest an absence of influence of refuge type on hydraulic
510 microhabitat selection. Nevertheless, our analysis of the influence of refuges could be complicated by
511 the influence of vegetated refuge such as woody debris on sampling technique efficiency (Thévenet &
512 Statzner, 1999) as well as difficulties to measure hydraulic conditions in refuges. In addition, the absence
513 of significant difference could be partly due to the lack of data with different type of refuges. We also
514 found that selection of substrate characteristics was weaker and less transferable across streams than
515 selection for hydraulics, as was observed earlier for fish and macroinvertebrates (Lamouroux, Dolédec &
516 Gayraud, 2004).

517 Fish size appears to be an important factor explaining habitat selection. Here, it was partly
518 accounted for by our use of specific size classes and its effects could deserve more detailed studies.
519 Indeed, for some species the models showed differences in average microhabitat selection between size
520 classes of the same species. *Barbus barbus* is a good example of the environmental gradient used by
521 individuals throughout their life. Smallest individuals (< 80 mm, mainly juveniles) select slower and
522 shallower habitat (velocity ~ 0.50 m.s⁻¹, depth ~ 0.25 m) than larger individuals (80 – 180 m), which
523 select themselves shallower habitat (velocity ~ 1 m.s⁻¹, depth ~ 0.50 m) than largest individuals
524 (velocity ~ 1 m.s⁻¹, depth ~ 0.75 m, size > 180 mm, mainly reproductive adult). Fish selection of deeper
525 habitats over time can be related to predator avoidance and the use of riverine habitats where they can
526 avoid high velocity variations. Nevertheless, considering the strong seasonal and annual variations in
527 fish length between and within species, our use of fixed fish size categories remains questionable. We
528 choose an arbitrary and uniform method for all species to define fish size classes, which was a
529 compromise between expert opinion and statistical difficulties. Further developments may consider
530 alternative groupings based on species life history strategies and traits (Schwartz 2016).

531 The strong variability explained by average multi-survey models adds evidence to the relevance
532 of generic habitat selection models, for many taxa, despite the variability of habitat selection across
533 surveys (e.g., Lamouroux *et al.*, 2013; Dixon & Vokoun, 2009). These results justify the application of
534 habitat selection models over river networks and catchments for appreciating the ecological impact of
535 flow management at large scale (e.g. Snelder, Booker & Lamouroux, 2011; Muñoz-Mas *et al.*, 2012).
536 The generality of habitat selection has also been explained by ecological strategies adapted to particular
537 hydraulic conditions, such as the association between opportunistic species and stressful habitat
538 conditions (Blanck, Tedesco & Lamouroux, 2007; Ayllón *et al.*, 2014). However, in many studies
539 involving fish or macroinvertebrates, it was found that habitat selection can vary a lot across space
540 (Lancaster & Downes, 2010) or time (Vilizzi, Copp & Roussel, 2004) due to several potential causes
541 such as variations in water quality or individual behavior. Our cross-validation values indicated also that
542 an average model does not systematically transfer well to an independent river. These variations could
543 be explained by the huge variation in river discharge (from 0.28 to 1490 m³.s⁻¹) and width (from 5 to 160
544 m), and the dominance of large rivers as the Rhône River and the Durance River regarding a few
545 number of small rivers as the Loup River and Paillons Rivers. These differences in river characteristics
546 were equally considered in our modeling approach and numerically limit the model transferability to small
547 rivers. All these studies illustrate the need to better examine the variations in habitat selection across
548 different regions or watersheds.

549 Part of the variability in microhabitat selection across surveys may be due to identification biases
550 of studied species. Indeed, recent studies revealed genetic differences within species identified using
551 taxonomic criteria in France such as *Gobio gobio*, *Barbatula barbatula* or *Esox lucius* (Kottelat & Freyhof,
552 2007; Denys, 2015). These differences suggest the presence of multiple species considered as a single

one in our study, and thus the development of microhabitat selection models at the genus level. Nevertheless, when species share taxonomic or functional traits, genus microhabitat selection models can be relevant as shown for macroinvertebrates genus such as *Glossosoma* or *Protonemura* (each grouping two species in Dolédec *et al.*, 2007). Similarly, microhabitat selection models developed for fish at the guild level showed common patterns of microhabitat selection and good transferability across populations and surveys (Lamouroux & Cattaneo, 2006, Martinez-Capel *et al.*, 2009). In addition, Chen & Olden (2018) showed that grouping species by guild based on their habitat use (i.e. species present in lower current velocity, for example *Cyprinus carpio* or *Gambusia affinis*) can provide relationships between flow and ecology that are transferable across rivers and basins, and thus relevant for catchment management.

Although low, our values of explained deviance (R^2_{MF}) and correlations between observed and fitted values (Spearman rho) indicated good model quality and good adjustment. Indeed, there is no absolute value to determine whether the model performs well and the significance of each evaluation metric must be discussed for each specific study (Guisan & Thuiller, 2005). In particular, when working at small scales such as the scale of microhabitats of a few square-meters, one cannot expect obtaining strong deterministic models of species abundance (see also Fladung, Scholten & Thiel, 2003). Our Bayesian approach of fit statistic (posterior predictive checks) clearly confirmed that the magnitude of our fits were typically those we could expect when considering the overdispersed character of fish abundance. These results showed the relevance of average microhabitat selection models despite apparent weak fit metrics.

Bivariate additive microhabitat selection, for water depth and water column current velocity, increased model performance. Indeed, for species presenting univariate selection for both hydraulic variables, model quality metrics increased in median from 0.04 to 0.06 for R^2_{MF} and from 0.30 to 0.38 for Spearman rho. These results confirm the need to further develop multivariate approach as described in previous studies on microhabitat and/or mesohabitat selection (e.g. Le Coarer, 2007; Dixon & Vokoun, 2009; Muñoz-Mas *et al.*, 2018). In contrast, for some species as *Barbus barbus*, the introduction of additive variable did not increase model performance and confirmed the assumption that in some cases univariate models perform better than multivariate ones (Millidine, Malcolm & Fryer, 2016). Several possibilities exist to investigate further the interactive effects of depth and velocity on microhabitat selection, such as describing habitats with the Froude number (Millidine *et al.* 2016), or using multivariate structural equation modeling (Leftwich *et al.* 1997), or Generalized Functional Response (Matthiopoulos *et al.* 2010). These approaches could improve model transferability across rivers.

Microhabitat selection models were often criticized in the literature for considering abundance as a proxy of habitat selection. Indeed, abundance or density are not always equal to selection and can be strongly influenced by density-dependent population dynamics and individual behavior (Van Horne, 1983; Lancaster & Downes, 2010; Lamouroux *et al.*, 2010). For example, for territorial species, the most

suitable microhabitat may be occupied by dominant individuals and thus correspond to low abundance (Rosenfeld, Beecher & Ptolemy, 2016). Alternatives to the use of abundance are to build selection models using presence or presence-absence data (e.g., Micheli-Campbell *et al.*, 2013; Guerra *et al.*, 2015), or to build bioenergetic models (Rosenfeld, Beecher & Ptolemy, 2016) that represent the energy gain or loss by organisms in a given microhabitat. Although encouraging, these bioenergetic models still deserve further field validation. Using other descriptors than abundance may raise other problems: for example, presence-absence data contain little information for species that have very low or very high occurrence. Our methodological approach considers overdispersion and has the advantage of taking into account the abundance information without giving too much importance to the samples with very high abundance. Specifically, our models treated differently gregarious species with schooling behavior, such as *Phoxinus phoxinus* (Garner, 1997), or less gregarious species such as *Barbus barbus*, and were well adapted to each condition.

Finally, our selection models considered the temporal and spatial variations of abundance over space (between microhabitats) and time (across surveys). However, they did not account for the temporal and spatial variations of the microhabitat characteristics themselves. Microhabitats where fish individuals were found have their own history influenced by temporal events such as strong variations in discharge (e.g., floods or hydropower releases, Kennard *et al.*, 2007, Capra *et al.*, 2017) or drying events (e.g. Pires, Beja & Magalhães, 2014). Similarly, these microhabitats belong to particular microhabitat spatial configurations- or distributions- that can be more or less heterogeneous (e.g., Martelo *et al.*, 2014) and more or less suitable for the different activities of the individuals over their dispersal distance (Radinger & Wolter, 2014). Ideally, these temporal and spatial variations in microhabitat characteristics around individuals should be described and included in microhabitat selection models. Our flexible mixed-effect methods could be further developed to integrate these aspects and provide more realistic, multi-scale models of complex habitat selection patterns.

DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting the findings of this study are available within the article. Any additional data requirements are available on request from the corresponding author.

REFERENCES

Ahmadi-Nedushan B., St-Hilaire A., Berube M., Robichaud E., Thiemonge N. & Bobee B. (2006) A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications* **22**, 503–523.

Alcaraz-Hernandez J.D., Munoz-Mas R., Martinez-Capel F., Garofano-Gomez V. & Vezza P. (2016) Generalized additive models to predict adult and young brown trout (*Salmo trutta* Linnaeus, 1758) densities in Mediterranean rivers. *Journal of Applied Ichthyology* **32**, 217–228.

Ayllón D., Nicola G.G., Parra I., Elvira B. & Almodóvar A. (2014). Spatio-temporal habitat selection shifts in brown trout populations under contrasting natural flow regimes. *Ecohydrology* **7**, 569–579.

Baker M.C. (2006) Differentiation of Mating Vocalizations in Birds: Acoustic Features in Mainland and Island Populations and Evidence of Habitat-Dependent Selection on Songs. *Ethology* **112**, 757–771.

Bates D., Maechler M., Bolker B. & Walker S. (2015) Fitting Linear Mixed-Effects Models using lme4. *Journal of statistical software* **67**, 1–48.

Blanck A. & Lamouroux N. (2007) Large-scale intraspecific variation in life-history traits of European freshwater fish. *Journal of Biogeography* **34**, 862–875.

Booker D.J. & Graynoth E. (2013) Relative influence of local and landscape-scale features on the density and habitat preferences of longfin and shortfin eels. *New Zealand Journal of Marine and Freshwater Research* **47**, 1–20.

Bunce R.G.H., Bogers M.M.B., Evans D., Halada L., Jongman R.H.G., Mucher C.A., *et al.* (2013) The significance of habitats as indicators of biodiversity and their links to species. *Ecological Indicators* **33**, 19–25.

Burnham K.P. & Anderson D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer, New-York.

Capra H., Plichard L., Bergé J., Pella H., Ovidio M., McNeil E., *et al.* (2017) Fish habitat selection in a large hydropowering river: Strong individual and temporal variations revealed by telemetry. *Science of The Total Environment* **578**, 109–120.

Chambert T., Rotella J.J. & Higgs M.D. (2014) Use of posterior predictive checks as an inferential tool for investigating individual heterogeneity in animal population vital rates. *Ecology and Evolution* **4**, 1389–1397.

Chen W. & Olden J.D. (2018) Evaluating transferability of flow–ecology relationships across space, time and taxonomy. *Freshwater Biology* **63**, 817–830.

Conallin J., Boegh E. & Jensen J.K. (2010) Instream physical habitat modeling types: an analysis as stream hydromorphological modeling tools for EU water resource managers. *International Journal of River Basin Management* **8**, 93–107.

- 651 Conallin J., Boegh E., Olsen M., Pedersen S., Dunbar M.J. & Jensen J.K. (2014) Daytime habitat
652 selection for juvenile parr brown trout (*Salmo trutta*) in small lowland streams. *Knowledge and*
653 *Management of Aquatic Ecosystems*, 09.
- 654 Costa M.R. da, Mattos T.M., Borges J.L. & Araujo F.G. (2013) Habitat preferences of common native
655 fishes in a tropical river in Southeastern Brazil. *Neotropical Ichthyology* **11**, 871–880.
- 656 Denys G. (2015) *Taxonomie intégrative des poissons d'eau douce de France métropolitaine*. Muséum
657 National d'Histoire Naturelle, Paris.
- 658 Dixon C.J. & Vokoun J.C. (2009). Burbot resource selection in small streams near the southern extent of
659 the species range. *Ecology of Freshwater Fish* **18**, 234–246.
- 660 Dolédec S., Lamouroux N., Fuchs U. & Méricoux S. (2007) Modeling the hydraulic preferences of
661 benthic macroinvertebrates in small European streams. *Freshwater Biology* **52**, 145–164.
- 662 Dunbar M.J., Alfredsen K. & Harby A. (2012) Hydraulic habitat modeling for setting environmental river
663 flow needs for salmonids. *Fisheries Management and Ecology* **19**, 500–517.
- 664 Fausch K.D., Torgersen C.E., Baxter C.V. & Li H.W. (2002) Landscapes to Riverscapes: Bridging the
665 Gap between Research and Conservation of Stream Fishes. *BioScience* **52**, 483–498.
- 666 Fisher J.T., Anholt B. & Volpe J.P. (2011) Body mass explains characteristic scales of habitat selection
667 in terrestrial mammals. *Ecology and Evolution* **1**, 517–528.
- 668 Fladung E., Scholten M. & Thiel R. (2003) Modeling the habitat preferences of preadult and adult fishes
669 on the shoreline of the large, lowland Elbe River. *Journal of Applied Ichthyology* **19**, 303–314.
- 670 Fretwell S.D. & Lucas H.L. (1969) On territorial behavior and other factors influencing habitat distribution
671 in birds. *Acta Biotheoretica* **19**, 16–36.
- 672 Fukuda S. (2011) Assessing the applicability of fuzzy neural networks for habitat preference evaluation
673 of Japanese medaka (*Oryzias latipes*). *Ecological Informatics* **6**, 286–295.
- 674 Garbe J. & Beeyers L. (2017) Modeling the impacts of a water trading scheme on freshwater habitats.
675 *Ecological Engineering* **105**, 284–295.
- 676 Garner P. (1997) Effects of variable discharge on the velocity use and shoaling behaviour of *Phoxinus*
677 *phoxinus*. *Journal of Fish Biology* **50**, 1214–1220.
- 678 Girard V., Monti D., Valade P., Lamouroux N., Mallet J.-P. & Grondin H. (2014) Hydraulic Preferences of
679 Shrimps and Fishes in Tropical Insular Rivers. *River Research and Applications* **30**, 766–779.
- 680 Goldstein H. (2003) *Multilevel statistical models*. John Wiley & Sons.
- 681 Gordon N.D., McMahon T.A. & Finlayson B.L. (1992) *Stream Hydrology: An Introduction for Ecologists*.
682 John Wiley & Sons Ltd, Chichester.

683 Gosselin M.-P., Maddock I. & Petts G. (2012) Mesohabitat use by brown trout (*Salmo trutta*) in a small
684 groundwater-dominated stream. *River Research and Applications* **28**, 390–401.

685 Grajeda L.M., Ivanescu A., Saito M., Crainiceanu C., Jaganath D., Gilman R.H., *et al.* (2016) Modeling
686 subject-specific childhood growth using linear mixed-effect models with cubic regression splines.
687 *Emerging Themes in Epidemiology* **13**.

688 Guerra Á., Hernández-Urcera J., Garci M.E., Sestelo M., Regueira M., González Á.F., *et al.* (2015)
689 Spawning habitat selection by *Octopus vulgaris*: New insights for a more effective management
690 of this resource. *Fisheries Research* **167**, 313–322.

691 Guisan A. & Thuiller W. (2005) Predicting species distribution: offering more than simple habitat models.
692 *Ecology Letters* **8**, 993–1009.

693 Guisan A. & Zimmermann N.E. (2000) Predictive habitat distribution models in ecology. *Ecological*
694 *Modeling* **135**, 147–186.

695 Hayes J.W., Hughes N.F. & Kelly L.H. (2007) Process-based modeling of invertebrate drift transport, net
696 energy intake and reach carrying capacity for drift-feeding salmonids. *Ecological Modeling* **207**,
697 171–188.

698 Ives A.R. (2015) For testing the significance of regression coefficients, go ahead and log-transform count
699 data. *Methods in Ecology and Evolution* **6**, 828–835.

700 Johnson D.H. (1980) The comparison of usage and availability measurements for evaluating resource
701 preference. *Ecology* **61**, 65–71.

702 Jowett I.G. & Davey A.J.H. (2007) A Comparison of Composite Habitat Suitability Indices and
703 Generalized Additive Models of Invertebrate Abundance and Fish Presence–Habitat Availability.
704 *Transactions of the American Fisheries Society* **136**, 428–444.

705 Kemp J.L., Harper D.M. & Crosa G.A. (1999) Use of ‘functional habitats’ to link ecology with morphology
706 and hydrology in river rehabilitation. *Aquatic Conservation: Marine and Freshwater Ecosystems*
707 **9**, 159–178.

708 Kennard M.J., Olden J.D., Arthington A.H., Pusey B.J. & Poff N.L. (2007) Multiscale effects of flow
709 regime and habitat and their interaction on fish assemblage structure in eastern Australia.
710 *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 1346–1359.

711 Kottelat M. & Freyhof J. (2007) *Handbook of European freshwater fishes*. Publications Kottelat.

712 Labonne J., Allouche S. & Gaudin P. (2003) Use of a generalised linear model to test habitat
713 preferences: the example of *Zingel asper*, an endemic endangered percoid of the River Rhône.
714 *Freshwater Biology* **48**, 687–697.

715 Lamouroux N., Capra H., Pouilly M. & Souchon Y. (1999) Fish habitat preferences in large streams of
716 southern France. *Freshwater Biology* **42**, 673–687.

717 Lamouroux N. & Cattaneo F. (2006) Fish assemblages and stream hydraulics: consistent relations
718 across spatial scales and regions. *River Research and Applications* **22**, 727–737.

719 Lamouroux N., Dolédec S. & Gayraud S. (2004) Biological traits of stream macroinvertebrate
720 communities: effects of microhabitat, reach, and basin filters. *Journal of the North American*
721 *Benthological Society* **23**, 449–466.

722 Lamouroux N., Mergoux S., Capra H., Dolédec S., Jowett I.G. & Statzner B. (2010) The generality of
723 abundance-environment relationships in microhabitats: A comment on Lancaster and Downes
724 (2009). *River Research and Applications* **26**, 915–920.

725 Lamouroux N., Mergoux S., Dolédec S. & Snelder T.H. (2013) Transferability of Hydraulic Preference
726 Models for Aquatic Macroinvertebrates. *River Research and Applications* **29**, 933–937.

727 Lancaster J. & Downes B.J. (2010) Linking the hydraulic world of individual organisms to ecological
728 processes: Putting ecology into ecohydraulics. *River Research and Applications* **26**, 385–403.

729 Le Coarer Y. (2007) Hydraulic signatures for ecological modeling at different scales. *Aquatic Ecology* **41**,
730 451–459.

731 Lehner B., Liermann C.R., Revenga C., Vörösmarty C., Fekete B., Crouzet P., *et al.* (2011) High-
732 resolution mapping of the world's reservoirs and dams for sustainable river-flow management.
733 *Frontiers in Ecology and the Environment* **9**, 494–502.

734 Magoulick, D.D., and Kobza, R.M. 2003. The role of refugia for fishes during drought: a review and
735 synthesis. *Freshw. Biol.* **48**(7): 1186–1198.

736 Martelo J., Grossman G.D., Porto M. & Filomena Magalhaes M. (2014) Habitat patchiness affects
737 distribution and microhabitat use of endangered Mira chub *Squalius torgalensis* (Actinopterygii,
738 Cypriniformes). *Hydrobiologia* **732**, 93–109.

739 Martinez-Capel F., García de Jalón D., Werenitzky D., Baeza D. & Rodilla-Alama M. (2009). Microhabitat
740 use by three endemic Iberian cyprinids in Mediterranean rivers (Tagus River Basin, Spain).
741 *Fisheries Management and Ecology* **16**, 52–60

742 McFadden D. (1974) The measurement of urban travel demand. *Journal of Public Economics* **3**, 303–
743 328.

744 Micheli-Campbell M.A., Campbell H.A., Connell M., Dwyer R.G. & Franklin C.E. (2013) Integrating
745 telemetry with a predictive model to assess habitat preferences and juvenile survival in an
746 endangered freshwater turtle. *Freshwater Biology* **58**, 2253–2263.

747 Millidine K.J., Malcolm I.A. & Fryer R.J. (2016) Assessing the transferability of hydraulic habitat models
748 for juvenile Atlantic salmon. *Ecological Indicators* **69**, 434–445.

749 Mouton A.M., Dillen A., Van den Neucker T., Buysse D., Stevens M. & Coeck J. (2012) Impact of
750 sampling efficiency on the performance of data-driven fish habitat models. *Ecological Modeling*
751 **245**, 94–102.

752 Muñoz-Mas R., Fukuda S., Portoles J. & Martinez-Capel F. (2018) Revisiting probabilistic neural
753 networks: a comparative study with support vector machines and the microhabitat suitability for
754 the Eastern Iberian chub (*Squalius valentinus*). *Ecological Informatics* **43**, 24–37.

755 Muñoz-Mas R., Martínez-Capel F., Alcaraz-Hernández J.D. & Mouton A.M. (2017) On species
756 distribution modeling, spatial scales and environmental flow assessment with Multi-Layer
757 Perceptron Ensembles: A case study on the redfin barbel (*Barbus haasi*; Mertens, 1925).
758 *Limnologia - Ecology and Management of Inland Waters* **62**, 161–172.

759 Muñoz-Mas R., Martinez-Capel F., Schneider M. & Mouton A.M. (2012) Assessment of brown trout
760 habitat suitability in the Jucar River Basin (SPAIN): Comparison of data-driven approaches with
761 fuzzy-logic models and univariate suitability curves. *Science of the Total Environment* **440**, 123–
762 131.

763 Nykanen M. & Huusko A. (2003) Size-related changes in habitat selection by larval grayling (*Thymallus*
764 *thymallus* L.). *Ecology of Freshwater Fish* **12**, 127–133.

765 Odum E. (1953) *Fundamentals of ecology*. W.B. Saunders Company, Philadelphia.

766 O'Hara R.B. & Kotze D.J. (2010) Do not log-transform count data. *Methods in Ecology and Evolution* **1**,
767 118–122.

768 Olden J.D., Konrad C.P., Melis T.S., Kennard M.J., Freeman M.C., Mims M.C., *et al.* (2014) Are large-
769 scale flow experiments informing the science and management of freshwater ecosystems?
770 *Frontiers in Ecology and the Environment* **12**, 176–185.

771 Pan H.Q. & Goldstein H. (1998) Multi-level repeated measures growth modeling using extended spline
772 functions. *Statistics in Medicine* **17**, 2755–2770.

773 Papadaki C., Soulis K., Ntoanidis L., Zogaris S., Dercas N. & Dimitriou E. (2017) Comparative
774 Assessment of Environmental Flow Estimation Methods in a Mediterranean Mountain River.
775 *Environmental Management* **60**, 280–292.

776 Pires D.F., Beja P. & Magalhães M.F. (2014) Out of Pools: Movement Patterns of Mediterranean Stream
777 Fish in Relation to Dry Season Refugia. *River Research and Applications*, n/a-n/a.

778 Pleydell D.R.J. & Chrétien S. (2010) Mixtures of GAMs for Habitat Suitability Analysis with
779 Overdispersed Presence/Absence Data. *Comput. Stat. Data Anal.* **54**(5), 1405–1418.

780 Poff N.L. (1997) Landscape Filters and Species Traits: Towards Mechanistic Understanding and
781 Prediction in Stream Ecology. *Journal of the North American Benthological Society* **16**, 391–409.

782 Poff N.L., Richter B.D., Arthington A.H., Bunn S.E., Naiman R.J., Kendy E., *et al.* (2010) The ecological
783 limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental
784 flow standards. *Freshwater Biology* **55**, 147–170.

785 Potts J.M. & Elith J. (2006) Comparing species abundance models. *Ecological Modeling* **199**, 153–163.

786 R Core Team (2017) *R: A language and environment for statistical computing*. R Foundation for
787 Statistical Computing, Vienna, Austria.

788 Radinger J., Kail J. & Wolter C. (2017) Differences among Expert Judgments of Fish Habitat Suitability
789 and Implications for River Management. *River Research and Applications* **33**, 538–547.

790 Radinger J. & Wolter C. (2014) Patterns and predictors of fish dispersal in rivers. *Fish and Fisheries* **15**,
791 456–473.

792 Riffart R., Carrel G., Le Coarer Y. & Fontez B.N. (2009) Spatio-temporal patterns of fish assemblages in
793 a large regulated alluvial river. *Freshwater Biology* **54**, 1544–1559.

794 Rosenfeld J., Beecher H. & Ptolemy R. (2016) Developing Bioenergetic-Based Habitat Suitability Curves
795 for Instream Flow Models. *North American Journal of Fisheries Management* **36**, 1205–1219.

796 Rosenfeld J.S. (2017) Developing flow–ecology relationships: Implications of nonlinear biological
797 responses for water management. *Freshwater Biology* **62**, 1305–1324.

798 Shiroyama R. & Yoshimura C. (2016) Assessing bluegill (*Lepomis macrochirus*) habitat suitability using
799 partial dependence function combined with classification approaches. *Ecological Informatics* **35**,
800 9–18.

801 Snelder T., Booker D. & Lamouroux N. (2011) A Method to Assess and Define Environmental Flow
802 Rules for Large Jurisdictional Regions 1. *JAWRA Journal of the American Water Resources*
803 *Association* **47**, 828–840.

804 Spearman C. (1904) The proof and measurement of association between two things. *The American*
805 *journal of psychology* **15**, 72–101.

806 Thévenet A. & Statzner B. (1999) Linking fluvial fish community to physical habitat in large woody debris:
807 sampling effort, accuracy and precision. *Archiv Fur Hydrobiologie* **145**, 57–77.

808 Tomsic C.A., Granata T.C., Murphy R.P. & Livchak C.J. (2007) Using a coupled eco-hydrodynamic
809 model to predict habitat for target species following dam removal. *Ecological Engineering* **30**,
810 215–230.

811 Udvardy M.F.D. (1959) Notes on the Ecological Concepts of Habitat, Biotope and Niche. *Ecology* **40**,
812 725–728.

813 Vadas R. & Orth D. (1993) A New Technique for Estimating the Abundance and Habitat Use of Stream
814 Fishes. *Journal of Freshwater Ecology* **8**, 305–317.

815 Van Horne B. (1983) Density as a misleading indicator of habitat quality. *The Journal of Wildlife*
816 *Management*, 893–901.

817 Vaudor L., Lamouroux N. & Olivier J.-M. (2011) Comparing distribution models for small samples of
818 overdispersed counts of freshwater fish. *Acta Oecologica* **37**, 170–178.

819 Vaudor L., Lamouroux N., Olivier J.-M. & Forcellini M. (2015) How sampling influences the statistical
820 power to detect changes in abundance: an application to river restoration. *Freshwater Biology* **60**,
821 1192–1207.

822 Vezza P., Parasiewicz P., Calles O., Spairani M. & Comoglio C. (2014) Modeling habitat requirements of
823 bullhead (*Cottus gobio*) in Alpine streams. *Aquatic Sciences* **76**, 1–15.

824 Vilizzi L., Copp G.H. & Roussel J.-M. (2004) Assessing variation in suitability curves and electivity
825 profiles in temporal studies of fish habitat use. *River Research and Applications* **20**, 605–618.

826 Vörösmarty C.J., McIntyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P., *et al.* (2010) Global
827 threats to human water security and river biodiversity. *Nature* **467**, 555–561.

828 Wagner T., Diefenbach D.R., Christensen S.A. & Norton A.S. (2011) Using multilevel models to quantify
829 heterogeneity in resource selection. *The Journal of Wildlife Management* **75**, 1788–1796.

830 Warton D.I. (2005) Many zeros does not mean zero inflation: comparing the goodness-of-fit of
831 parametric models to multivariate abundance data. *Environmetrics* **16**, 275–289.

832 Warton D.I., Wright S.T. & Wang Y. (2012) Distance-based multivariate analyses confound location and
833 dispersion effects. *Methods in Ecology and Evolution* **3**, 89–101.

834 Wentworth C.K. (1922) A Scale of Grade and Class Terms for Clastic Sediments. *The Journal of*
835 *Geology* **30**, 377–392.

836 Wisz M.S., Hijmans R.J., Li J., Peterson A.T., Graham C.H., Guisan A., *et al.* (2008) Effects of sample
837 size on the performance of species distribution models. *Diversity and Distributions* **14**, 763–773.

838 Wood S.N. (2006) *Generalized Additive Models : An Introduction with R*. Chapman and Hall/CRC.

839 Zuur A.F., Saveliev A. & Ieno E.N. (2014) *A Beginner's Guide to Generalised Additive Mixed Models with*
840 *R*, Highland Statistics Ltd.

Table 1: River characteristics: number of reaches sampled, number of surveys, median number of microhabitats sampled during each survey, mean daily discharge [m³.s⁻¹] average by sampling date and for the entire reach, average width of the reach [m]

River	Nb of reaches	Nb of surveys	Median Nb of microhabitats per survey	Mean daily discharge			Average width
				Average by sampling date		Reach	Reach
				Minimum	Maximum		
Ain	1	9	51	20.80	155.00	125	60
Ardèche	1	12	50	0.98	33.85	24	30
Drôme	1	2	37	13.10	27.20	7	20
Durance	18	1 to 12	15	1.01 to 78.50	1.01 to 270.00	6 to 20	15 to 100
Garonne	1	2	63	52.20	59.50	200	100
Le Loup	2	1	100	0.49	0.83	7	10
Les Paillons	6	1	30	0.16	0.24	0.28	5
Loire	5	1	15	10.50	19.00	50	75
Rhône	4	1 to 7	50	609.00	855.00	1030 to 1490	120 to 160

842

843

Table 2: Species codes, size classes and characteristics. Total occurrence is the number of microhabitats where the size class occurred, among the total number of microhabitats sampled during the surveys where the size class occurred. Underlined specific size classes were considered for multivariate models.

Family	Species	Code	Size class	Total Abundance	Total Occurrence	Total Nb of microhabitats	Nb of Surveys	% of data with substratum	% of microhabitats with refuges	
									Absence	Presence
Anguillidae										
	<i>Anguilla anguilla</i>	AnA	cl23	177	84	250	4	-	0.73	0.27
			cl4	554	273	761	16	0.37	0.45	0.55
Balitoridae										
	<i>Barbatula barbatula</i>	BaBa	cl12	14661	1124	2359	98	0.41	0.72	0.28
Centrarchidae										
	<i>Lepomis gibbosus</i>	LeG	cl123	974	159	482	14	0.49	0.43	0.57
Cobitidae										
	<i>Cobitis bilineata</i>	CoB	cl12	518	127	396	19	0.29	0.87	0.13
Cottidae										
	<i>Cottus gobio</i>	CoG	cl12	200	77	312	11	0.46	0.56	0.44
Cyprinidae										
	<i>Alburnoides bipunctatus</i>	AIB	cl1	8506	831	1828	84	0.38	0.66	0.34
			cl23	3349	512	1370	60	0.40	0.68	0.32
	<i>Alburnus alburnus</i>	AIA	cl1	1918	193	811	22	0.48	0.53	0.47
			cl23	895	124	451	14	0.48	0.62	0.38
	<i>Barbus barbus</i>	BaBu	cl1	3848	691	1687	69	0.41	0.67	0.33
			cl2	2393	528	1454	63	0.41	0.60	0.40
			cl34	1836	287	1176	37	0.48	0.54	0.46
	<i>Barbus meridionalis</i>	BaM	cl1	235	31	199	3	-	1.00	-
			cl23	607	155	364	4	-	0.57	0.43
	<i>Blicca bjoerkna</i>	BIB	cl12	360	44	219	8	0.46	0.58	0.42
	<i>Chondrostoma nasus</i>	ChN	cl12	925	136	412	13	0.49	0.71	0.29
			cl34	298	67	465	9	0.48	0.74	0.26
	<i>Gobio gobio</i>	GoG	cl1	4580	643	1674	59	0.43	0.67	0.33
			cl23	2376	596	1770	60	0.45	0.61	0.39
	<i>Leuciscus leuciscus</i>	LeL	cl123	317	81	371	12	0.50	0.35	0.65
	<i>Parachondrostoma toxosoma</i>	PaT	cl1	1195	136	534	16	0.44	0.74	0.26
			cl23	338	85	423	12	0.48	0.65	0.35
	<i>Phoxinus phoxinus</i>	PhP	cl12	11721	844	1854	71	0.37	0.71	0.29
	<i>Rhodeus amarus</i>	RhA	cl12	66	13	37	3	0.50	-	1.00
	<i>Rutilus rutilus</i>	RuR	cl1	1844	145	558	20	0.49	0.51	0.49
			cl2	1373	148	505	16	0.48	0.36	0.64
			cl34	188	39	211	6	0.34	0.42	0.58
	<i>Squalius cephalus</i>	SqC	cl1	8065	906	2272	88	0.42	0.62	0.38
			cl2	3956	727	1854	79	0.43	0.51	0.49
			cl34	1129	341	1238	46	0.48	0.46	0.54

Table 2. Continued

Family	Species	Code	Size class	Total Abundance	Total Occurrence	Total Nb of microhabitats	Nb of Surveys	% of data with substratum	% of microhabitats with refuges	
									Absence	Presence
Cyprinidae	<i>Telestes soufia</i>	TeS	cl1	5231	625	1481	59	0.36	0.66	0.34
			cl2	2857	517	1644	60	0.37	0.60	0.40
Percidae	<i>Perca fluviatilis</i>	PeF	cl12	188	72	419	12	0.49	0.43	0.57
	<i>Zingel asper</i>	ZiA	cl12	29	18	196	3	0.50	1.00	-
Salmonidae	<i>Salmo trutta fario</i>	SaT	cl12	308	122	380	7	0.33	0.69	0.31
			cl34	87	34	93	3	-	-	1.00

844

845

Table 3: Univariate model fits for microhabitat characteristics without refuge × specific size classes for specific size class showing minimum magnitude. Δ QAIC: difference with the lowest QAIC among models M1-M4, McFadden's R^2 (R^2 MF) observed with the dataset (Obs) and the R^2 MF observed in the 95-percentile of the posterior predictive checks simulations (Simu), the difference between the R^2 MF of M1 and M4 (ΔR^2 MF), the relative R^2 MF of M2 compared with M4 (RR^2 MF), and similar statistics for Spearman rho. The values were ordered of decreasing ΔR^2 MF. Underlined specific size classes were considered for multivariate models. Bold numbers represent the selected model according to an AIC's selection.

Habitat	Species code	Class	Δ QAIC				McFadden's R^2				ΔR^2 MF	RR^2 MF	Spearman rho				Δ rho	Rrho				
			M1	M2	M3	M4	M1		M2				M4		M1				M2		M4	
							Obs	Simu	Obs	Simu			Obs	Simu	Obs	Simu			Obs	Simu	Obs	Simu
Water depth																						
	BaBa	cl12	681	22	2	-	0.03	0.09	0.12	0.16	0.12	0.19	0.09	0.96	0.37	0.62	0.67	0.71	0.69	0.72	0.32	0.93
	CoG	cl12	34	-	3	8	-	-	0.08	0.20	0.09	0.24	0.09	0.94	-0.10	0.35	0.29	0.52	0.31	0.73	0.41	0.95
	BaM	cl23	29	-	2	5	0.03	0.06	0.09	0.15	0.10	0.26	0.08	0.87	0.27	0.51	0.42	0.64	0.43	0.68	0.16	0.91
	<u>BaBu</u>	<u>cl1</u>	178	59	-	5	0.03	0.05	0.07	0.12	0.09	0.21	0.06	0.65	0.27	0.51	0.49	0.61	0.52	0.61	0.25	0.87
	<u>CoB</u>	<u>cl12</u>	25	-	3	9	-	-	0.05	0.09	0.06	0.08	0.06	0.95	-0.12	0.32	0.22	0.43	0.26	0.45	0.39	0.89
	<u>GoG</u>	<u>cl23</u>	111	5	5	-	0.01	0.03	0.05	0.07	0.06	0.08	0.04	0.82	0.07	0.37	0.24	0.57	0.31	0.64	0.24	0.73
	<u>BaBu</u>	<u>cl34</u>	41	-	3	2	0.05	0.05	0.08	0.09	0.09	0.16	0.04	0.88	0.20	0.38	0.32	0.51	0.37	0.49	0.17	0.69
	<u>AlA</u>	<u>cl23</u>	22	19	5	-	0.01	0.02	0.02	0.06	0.05	0.14	0.04	0.21	0.22	0.34	0.27	0.40	0.34	0.47	0.13	0.42
	<u>BaBu</u>	<u>cl2</u>	91	4	-	2	0.02	0.04	0.06	0.07	0.06	0.15	0.04	0.82	0.27	0.43	0.38	0.53	0.40	0.74	0.13	0.83
	<u>PhP</u>	<u>cl12</u>	208	66	1	-	0.02	0.05	0.05	0.08	0.06	0.08	0.04	0.67	0.26	0.44	0.46	0.51	0.52	0.54	0.26	0.75
	SqC	cl34	11	2	-	8	0.02	0.03	0.06	0.07	0.06	0.17	0.04	0.84	0.09	0.31	0.20	0.43	0.27	0.41	0.18	0.64
	<u>AlB</u>	<u>cl23</u>	71	2	-	8	0.02	0.04	0.04	0.07	0.05	0.08	0.03	0.85	0.30	0.43	0.40	0.53	0.45	0.56	0.15	0.71
	<u>AnA</u>	<u>cl4</u>	12	6	-	5	0.10	0.11	0.11	0.13	0.13	0.20	0.03	0.48	0.37	0.57	0.41	0.63	0.48	0.78	0.11	0.38
	<u>TeS</u>	<u>cl2</u>	62	18	6	-	0.02	0.02	0.04	0.05	0.05	0.17	0.03	0.57	0.17	0.36	0.28	0.44	0.34	0.44	0.17	0.63
	<u>GoG</u>	<u>cl1</u>	83	18	-	8	0.02	0.05	0.04	0.08	0.05	0.18	0.03	0.76	0.21	0.41	0.37	0.47	0.41	0.57	0.19	0.80
	<u>TeS</u>	<u>cl1</u>	98	10	-	3	0.01	0.03	0.03	0.06	0.03	0.06	0.03	0.82	0.10	0.41	0.32	0.46	0.37	0.47	0.27	0.80
	<u>SqC</u>	<u>cl2</u>	32	7	2	-	0.02	0.02	0.03	0.05	0.04	0.06	0.03	0.56	0.06	0.34	0.18	0.40	0.26	0.40	0.20	0.62
	<u>SqC</u>	<u>cl1</u>	111	30	2	-	0.01	0.02	0.03	0.05	0.03	0.12	0.03	0.68	-0.04	0.31	0.21	0.38	0.25	0.41	0.30	0.86
	<u>AlA</u>	<u>cl1</u>	14	9	13	-	0.02	0.04	0.03	0.04	0.04	0.08	0.02	0.30	0.16	0.33	0.18	0.34	0.35	0.41	0.19	0.08
	<u>PaT</u>	<u>cl1</u>	13	15	-	4	0.03	0.05	0.03	0.12	0.05	0.11	0.02	0.15	0.18	0.35	0.22	0.38	0.33	0.41	0.15	0.25
Current velocity																						
	PeF	cl12	23	-	-	5	-	0.06	0.14	0.24	0.16	0.29	0.16	0.85	-0.25	0.40	0.22	0.49	0.29	0.52	0.54	0.88
	BaM	cl1	29	1	-	4	0.05	0.10	0.16	0.25	0.19	0.36	0.15	0.82	0.05	0.52	0.34	0.52	0.37	0.54	0.33	0.88
	<u>CoB</u>	<u>cl12</u>	118	-	2	3	-	-	0.12	0.19	0.13	0.15	0.13	0.94	-0.12	0.37	0.47	0.60	0.49	0.58	0.61	0.97
	LeG	cl123	48	-	2	6	-	0.02	0.08	0.15	0.09	0.23	0.09	0.93	-0.22	0.35	0.37	0.54	0.40	0.64	0.62	0.96

Table 3. Continued

Habitat	Species code	Class	Δ QAIC				McFadden's R^2						ΔR^2 MF	RR ² MF	Spearman rho				Δ rho		Rrho	
			M1	M2	M3	M4	M1		M2		M4				M1	Simu	M2	Simu	M4	Simu		
							Obs	Simu	Obs	Simu	Obs	Simu										Obs
Current velocity																						
	BIB	cl12	34	-	2	6	0.01	0.05	0.09	0.12	0.10	0.14	0.09	0.91	0.12	0.38	0.36	0.53	0.41	0.52	0.29	0.82
	RuR	cl1	49	-	-	1	0.01	0.04	0.08	0.13	0.09	0.13	0.08	0.86	0.03	0.39	0.40	0.53	0.40	0.55	0.37	0.98
	BaBu	cl2	112	52	14	-	0.02	0.04	0.05	0.10	0.08	0.10	0.05	0.51	0.27	0.46	0.42	0.53	0.47	0.57	0.20	0.76
	SqC	cl1	244	26	5	-	0.01	0.03	0.05	0.09	0.06	0.08	0.05	0.86	-0.04	0.37	0.38	0.49	0.42	0.51	0.46	0.93
	GoG	cl1	192	39	-	1	0.02	0.05	0.06	0.09	0.07	0.11	0.05	0.75	0.21	0.42	0.39	0.48	0.42	0.52	0.21	0.85
	SqC	cl2	96	14	-	2	0.02	0.02	0.05	0.08	0.05	0.06	0.04	0.79	0.06	0.35	0.23	0.42	0.29	0.43	0.23	0.77
	BaBu	cl34	31	22	7	-	0.05	0.04	0.06	0.08	0.08	0.13	0.04	0.28	0.20	0.35	0.35	0.42	0.39	0.48	0.18	0.78
	AIA	cl1	38	5	8	-	0.02	0.04	0.04	0.07	0.05	0.09	0.04	0.72	0.16	0.34	0.23	0.39	0.33	0.44	0.17	0.44
	AIB	cl23	80	44	14	-	0.02	0.04	0.03	0.07	0.05	0.13	0.04	0.36	0.30	0.43	0.42	0.50	0.46	0.56	0.16	0.74
	PhP	cl12	184	44	4	-	0.02	0.05	0.05	0.08	0.06	0.08	0.04	0.71	0.26	0.45	0.44	0.49	0.52	0.54	0.26	0.69
	AIA	cl23	12	5	4	-	0.01	0.02	0.03	0.05	0.05	0.10	0.04	0.43	0.22	0.34	0.24	0.40	0.35	0.46	0.13	0.19
	AnA	cl4	14	9	-	4	0.10	0.10	0.11	0.14	0.13	0.18	0.03	0.38	0.37	0.57	0.41	0.60	0.46	0.60	0.09	0.44
	PaT	cl1	25	-	3	7	0.03	0.05	0.06	0.09	0.06	0.11	0.03	0.91	0.18	0.36	0.32	0.40	0.36	0.42	0.18	0.78
	TeS	cl1	114	10	11	-	0.01	0.03	0.03	0.07	0.04	0.06	0.03	0.84	0.10	0.41	0.29	0.47	0.36	0.50	0.26	0.73
	ChN	cl12	12	3	-	5	0.02	0.07	0.04	0.12	0.05	0.13	0.03	0.64	0.26	0.46	0.34	0.55	0.37	0.53	0.11	0.74
	GoG	cl23	59	20	-	2	0.01	0.02	0.03	0.04	0.03	0.06	0.02	0.62	0.07	0.34	0.13	0.36	0.23	0.41	0.16	0.41
	BaBu	cl1	68	46	4	-	0.03	0.05	0.04	0.06	0.05	0.08	0.02	0.33	0.27	0.47	0.32	0.50	0.41	0.55	0.14	0.39
	TeS	cl2	35	28	-	1	0.02	0.03	0.02	0.03	0.04	0.07	0.02	0.22	0.17	0.36	0.14	0.35	0.29	0.42	0.12	-0.24
Dominant substratum																						
	PeF	cl12	34	-	4	3	-	0.06	0.14	0.28	0.17	0.26	0.17	0.84	-0.23	0.41	0.18	0.44	0.28	0.46	0.51	0.80
	AIA	cl1	28	8	-	-	0.01	0.03	0.03	0.08	0.05	0.11	0.05	0.59	0.13	0.32	0.26	0.38	0.32	0.42	0.19	0.69
	SqC	cl34	11	10	8	-	0.02	0.03	0.03	0.06	0.07	0.08	0.05	0.15	0.06	0.30	0.09	0.32	0.20	0.37	0.14	0.18
	LeG	cl123	17	-	3	9	-	0.01	0.03	0.09	0.03	0.10	0.03	0.97	-0.22	0.30	0.13	0.34	0.16	0.38	0.38	0.92
	RuR	cl1	13	-	4	8	0.01	0.03	0.03	0.09	0.03	0.09	0.03	0.88	0.05	0.32	0.21	0.38	0.21	0.39	0.17	0.96
	AIB	cl23	36	22	-	6	0.02	0.06	0.03	0.07	0.04	0.16	0.02	0.39	0.31	0.45	0.37	0.47	0.42	0.56	0.11	0.52
	TeS	cl2	15	12	-	4	0.02	0.03	0.02	0.06	0.04	0.22	0.02	0.24	0.12	0.35	0.19	0.38	0.24	0.45	0.12	0.61
Substratum heterogeneity																						
	PeF	cl12	10	-	4	9	-	0.59	0.08	0.22	0.08	0.24	0.08	0.94	-0.23	0.28	0.10	0.35	0.18	0.38	0.41	0.81
	SqC	cl2	24	17	7	-	0.02	0.04	0.03	0.05	0.04	0.09	0.03	0.28	0.06	0.37	0.05	0.37	0.19	0.41	0.13	-0.10

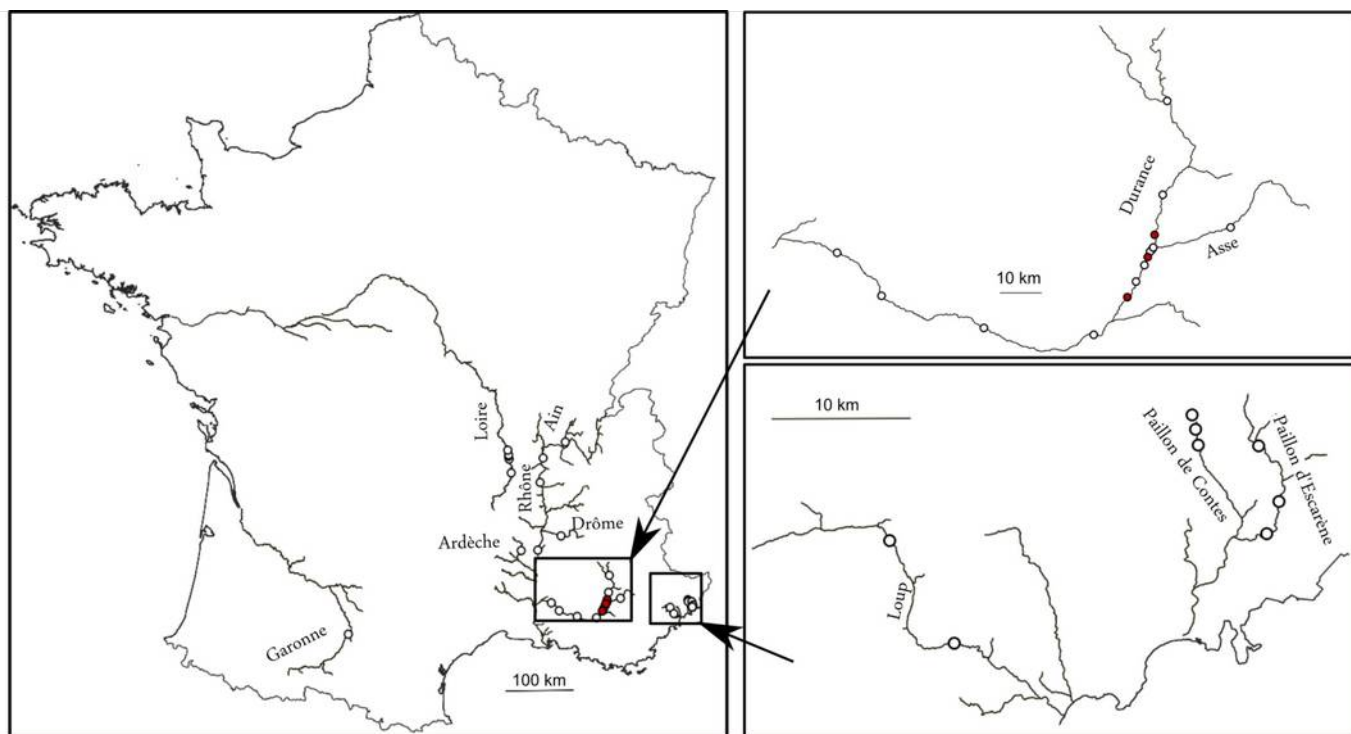


Figure 1: Location of sampling sites (points). Red points indicate location of two sites closely sampled.

848

849

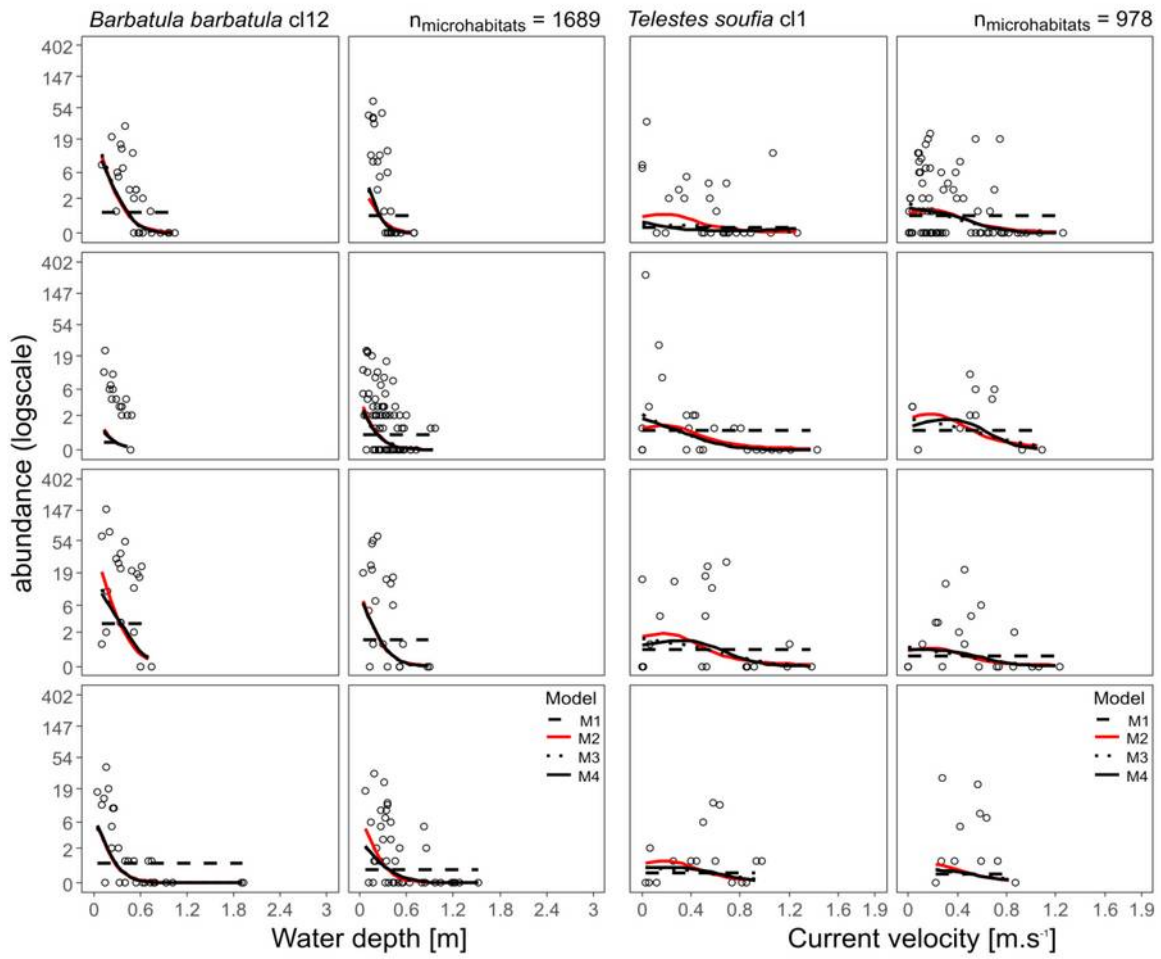


Figure 2: Observed microhabitat abundance (empty points) for conditional predictions of our four models (M1: black dashed lines, M2: red solid lines, M3: black dotted lines, M4: black solid lines) for the eight most abundant surveys for *Barbatula barbatula* cl12 and water depth (left) and for *Telestes soufia* cl1 and current velocity habitat (right). Each frame represents a survey. Because of their similarities (i.e. adjusted intercept and habitat selection shape by surveys), M3 and M4 lines are often superimposed on the graphs. Species codes are in Table 2.

850

851

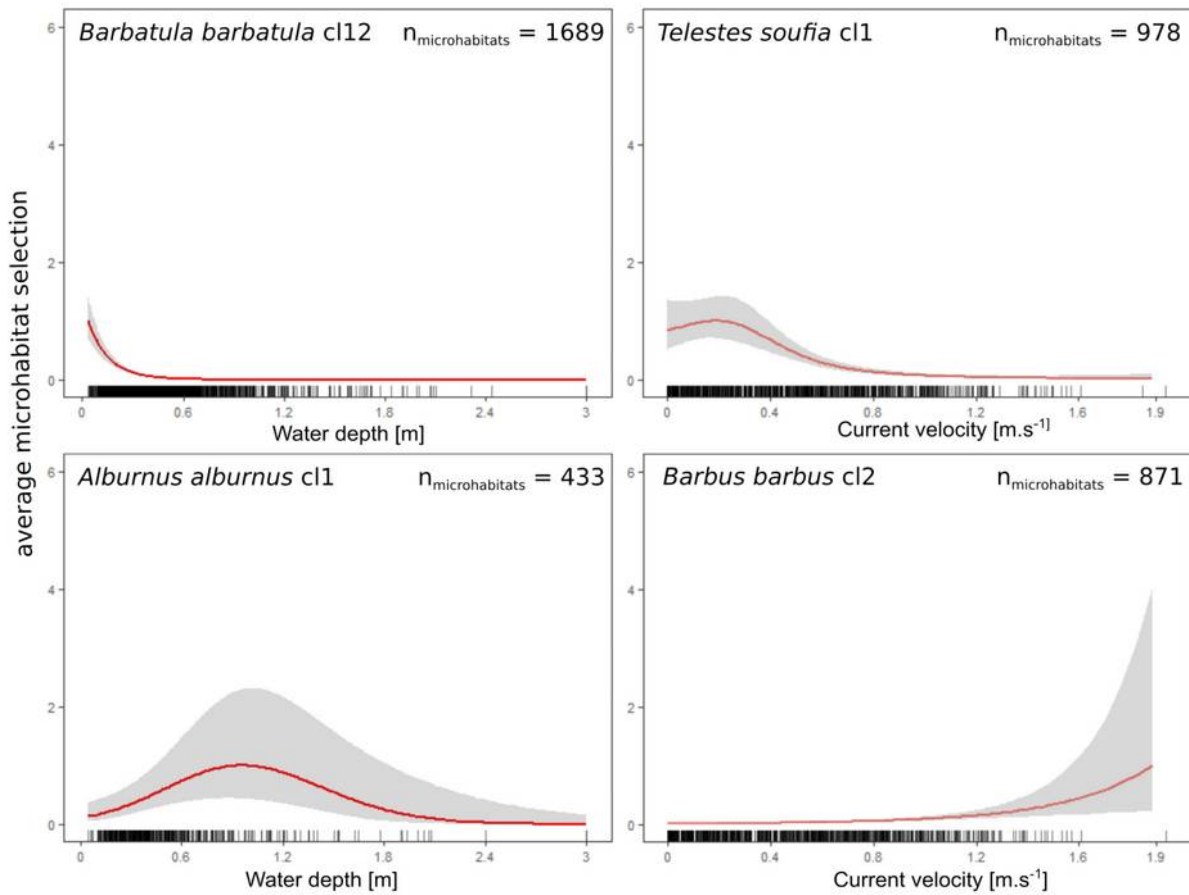


Figure 3: Marginal predictions of model M2 (red lines) with its 95-percent confidence interval (grey areas) for *Barbatula barbatula* cl12 and *Alburnus alburnus* cl1 in relation to water depth (left) and *Telestes soufia* cl1 and *Barbus barbus* cl2 in relation to current velocity (right). The Y-axis represents the average microhabitat selection corresponding to the ratio of the predicted abundance and the maximum predicted abundance of the “average” model M2. The dashes along the x-axis indicate the distribution of measurements of fish presence in the model. Species codes are from Table 2.

852

853

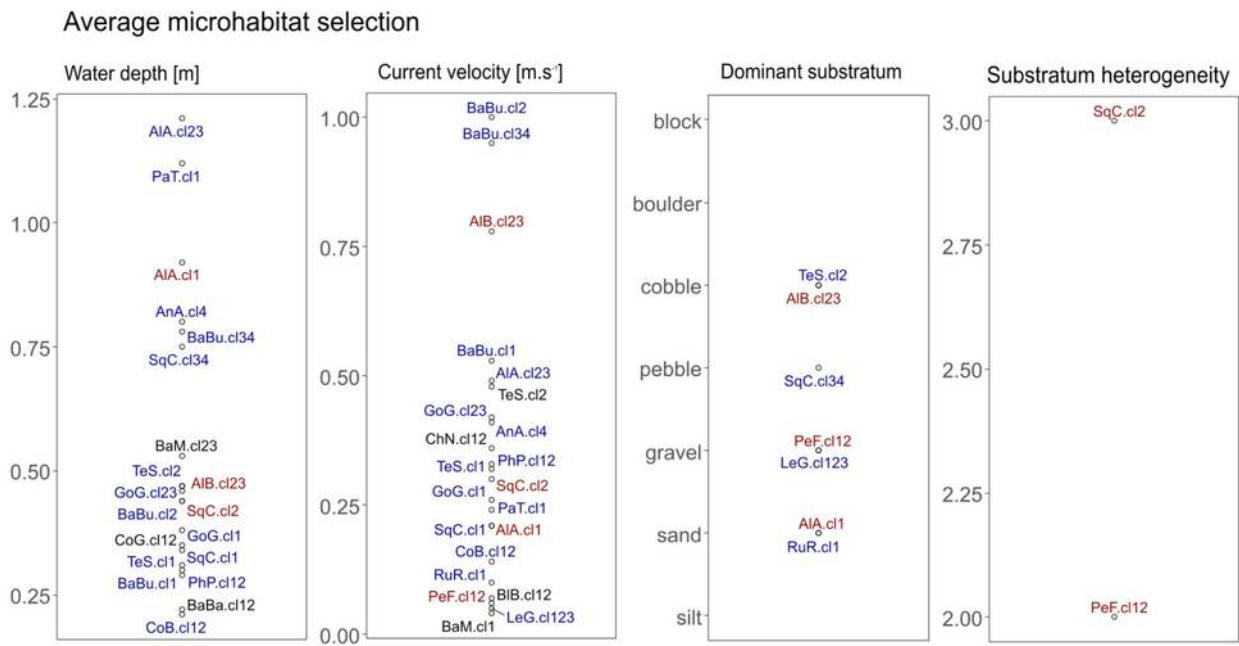


Figure 4: Average microhabitat selection corresponding to the average value of the marginal component of M2 (see examples in Fig. 2) over the range of habitat characteristics observed in our dataset. Codes of species and size classes are from Table 2. Blue codes represent specific size classes showing average microhabitat selection for two microhabitat types, red codes for three microhabitat types.

854

855

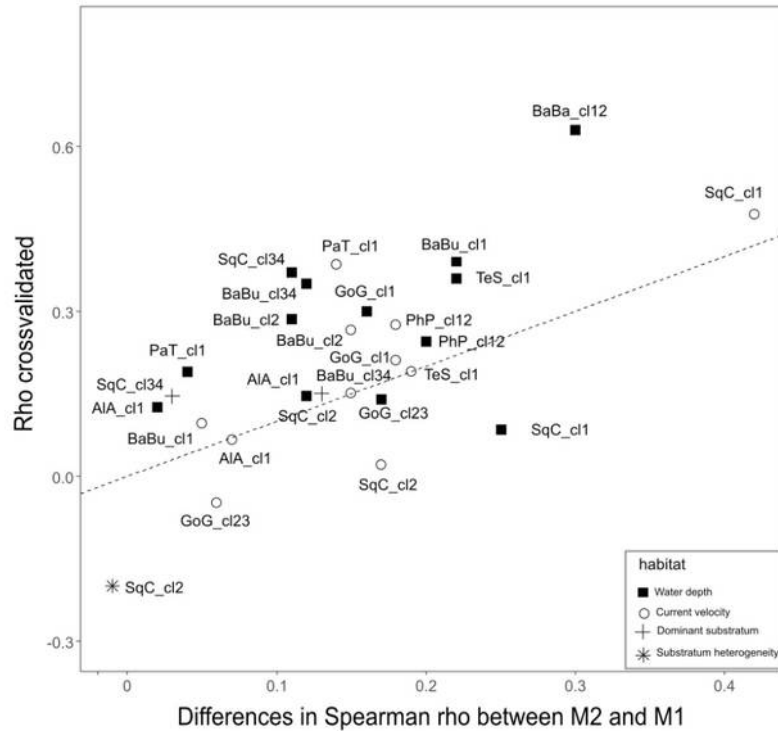


Figure 5: Distribution of Spearman rho values relating the observed ranks of abundance in surveys with predicted ranks (obtained for model M2 during our leave_one_river_out cross validations). These statistics indicate the potential of our "average" selection model M2 to predict blindly abundance ranks in an independent river. These cross-validations rho values are shown for the different specific size classes, ordered on the x-axis according to the differences of Spearman rho values between the full M2 and full M1. These differences indicate how M2 fits abundance ranks within surveys. Points indicate median values obtained across all the rivers tested for the specific size class.

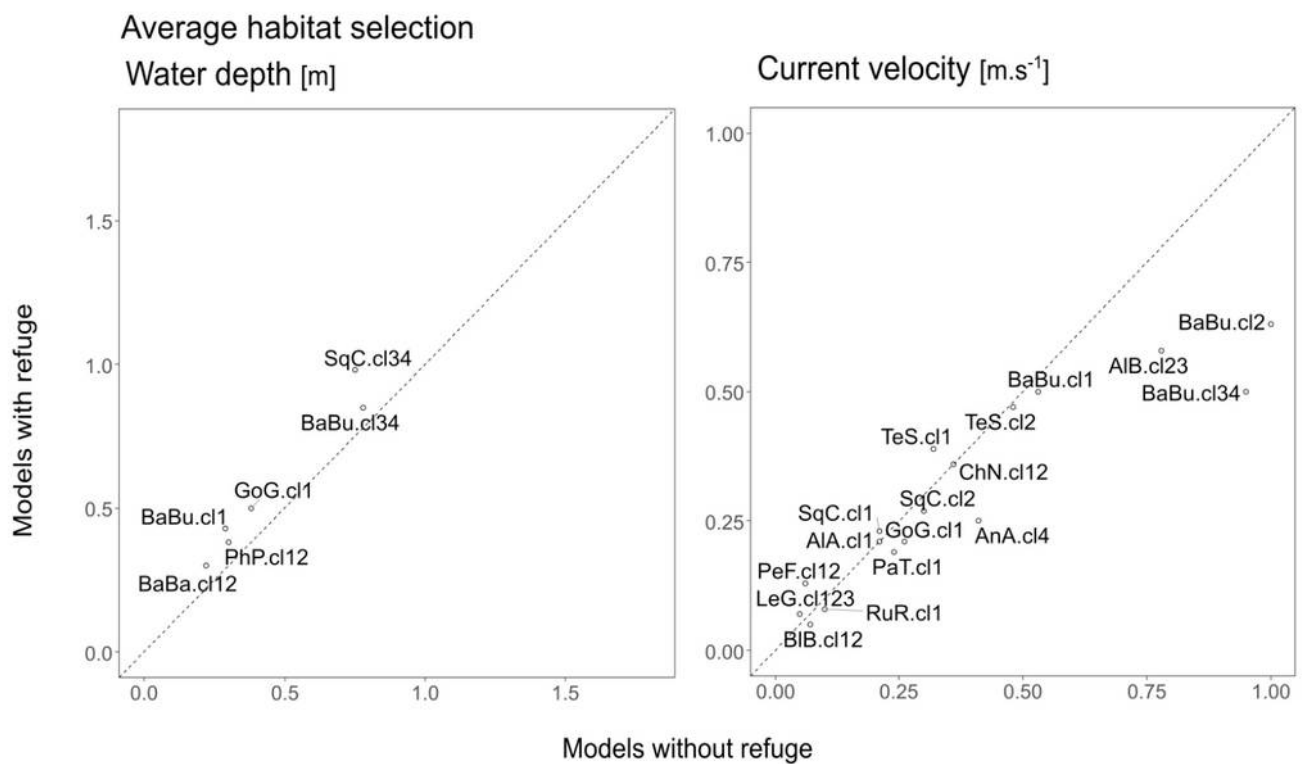


Figure 6: Comparison of the average microhabitat selected (defined as in Fig. 3) between models M2 developed without (X-axis) and with (Y-axis) refuge. Dashed lines represent $y=x$, corresponding to a similar microhabitat selection with and without refuge.

857

858

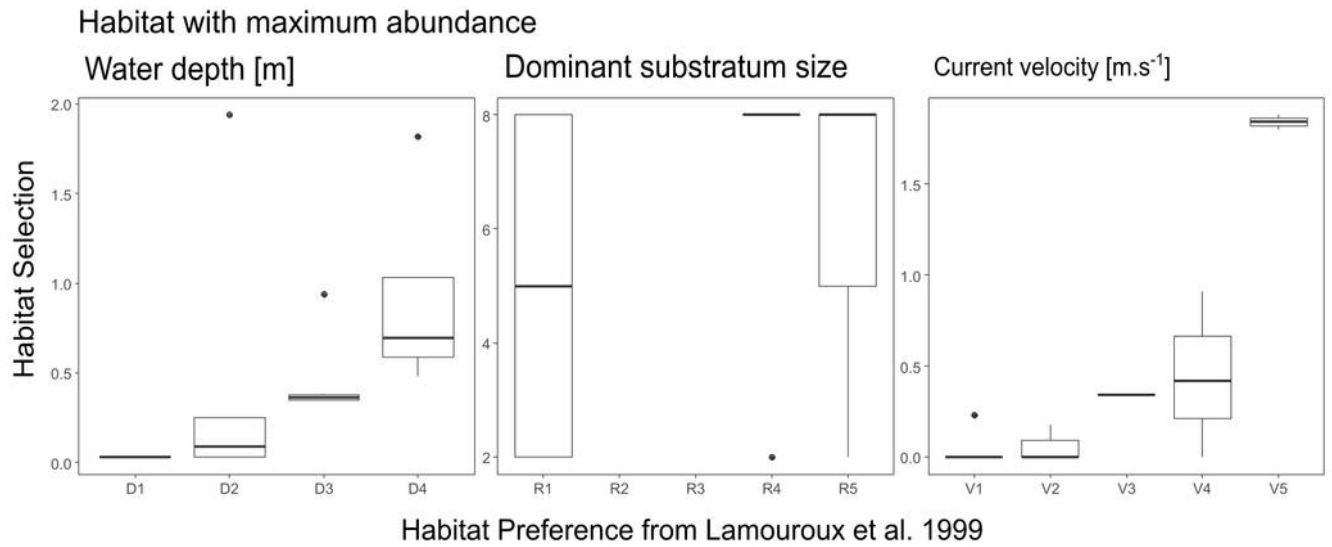


Figure 7: Comparison of habitat characteristics corresponding to maximum abundance values, between our models and those of Lamouroux *et al.* (1999). This comparison is made using the microhabitat classes from Lamouroux *et al.* (1999): **Water depth:** 0-0.2 m (D1); 0.2-0.4 m (D2); 0.4-0.8 m (D3); >0.8m (D4); **Dominant substratum:** 0-0.016 m (R1); 0.016-0.064 m (R2); 0.064-0.256 m (R3); >0.256 m (R4); large bedrocks (R5); **Current velocity** (i.e. water column water current velocity): 0-0.05 m.s⁻¹ (V1); 0.05-0.2 m.s⁻¹ (V2); 0.2-0.4 m.s⁻¹ (V3); 0.4-0.5 m.s⁻¹ (V4); >0.8 m.s⁻¹ (V5). Boxplots show the median, quartiles, 95% confidence intervals and extreme values across the different specific size classes.

859

860

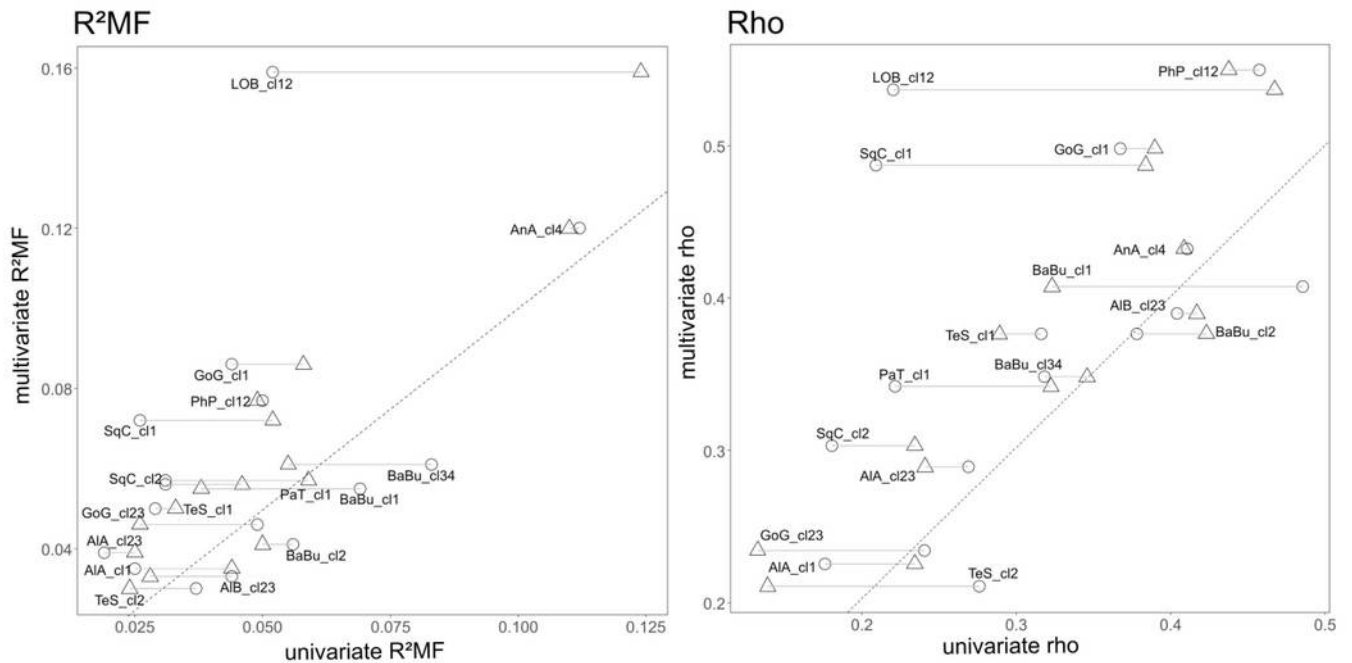


Figure 8: Comparison of univariate (X-axis) and multivariate (Y-axis) performance of model M2, indicated by McFadden's R^2_{MF} (left) and conditional Spearman's ρ (right). Circles and triangles represent respectively the performance of univariate models associated with water depth and current velocity. Grey segments link the performance for water depth and current velocity for a given specific class. Dashed lines represent the $Y=X$ axis.