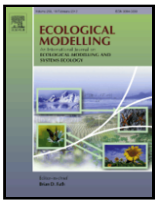
Ecological Modelling xxx (xxxx) 109905

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



A new look at habitat suitability curves through functional data analysis

Jeremie Boudreault[a, b, *], Normand E Bergeron[a, b], Andre St-Hilaire[a, b, c], Fateh Chebana[a, b]

[a] *Institut National de la Recherche Scientifique – Centre Eau Terre Environnement, 490 de la Couronne, Québec (QC) G1K 9A9, Canada*
[b] *Laboratoire d'Analyse et de Modélisation des Habitats Aquatiques (LAMHA), 490 de la Couronne, Québec (QC) G1K 9A9, Canada*
[c] *Canadian Rivers Institute, University of New Brunswick, P.O. Box 4400, Fredericton (NB) E3B 5A3, Canada*

ARTICLE INFO

Keywords:
Ecological modelling
Habitat suitability curves
Functional data analysis
Functional regression model
Fish habitat
juvenile Atlantic salmon

ABSTRACT

Habitat suitability curves (HSC) synthesize the preference of a species for important habitat variables and are, therefore, key components of various fish habitat models. However, HSC are developed at large scales (e.g. river or regional scales) that do not consider the differences that exist in available habitat conditions at smaller scales. To address this problem, a new look at HSC is taken through functional data analysis (FDA). It is an appropriate framework adapted for HSC construction because in FDA, each observation is a curve or a function. To illustrate the potential of FDA for HSC, a dataset of Atlantic salmon (*Salmo salar*) parr density and habitat variables constructed on two rivers was used. Functional regression models (FRM) were built to predict site-specific HSC based on the available habitat conditions for three salmon parr habitat variables: water depth, mean flow velocity and median substrate size. FRM explained a greater proportion of the variation in site-specific HSC (respectively 38.0%, 53.3% and 45.5% for depth, substrate size and velocity) compared to traditional HSC developed at the scale of each river or regionally that poorly fitted site-specific HSC. When HSC were aggregated into habitat suitability indices (HSI), weak relationships were found between HSI and parr density ($R^2 < 5\%$) for all models (traditional HSC and FRM). This study demonstrates that FDA is an innovative framework that can be used to predict more representative site-specific HSC adapted to differences in local available habitat. The results suggested that its potential should be further exploited in habitat modelling.

1. Introduction

Habitat suitability curves (HSC) are used to model habitat preferences of a given species (DeGraaf and Bain, 1986; Morantz *et al.*, 1987; Heggenes, 1996). HSC are then used as a key component of habitat simulation models (Bovee, 1982; Leclerc *et al.*, 1995). In river ecology, for example, HSC are combined with outputs from a hydraulic model in tools such as PHABSIM (Bovee, 1982; Bovee *et al.*, 1998), HABSIM (Parasiewicz, 2001) or CASiMiR (Jorde *et al.*, 2001). Habitat models are frequently used in river management, allowing for the estimation of the potential fish production of a river, to predict the effect of changes in flow regime on fish habitat or to simulate stream restoration measures (e.g. Leclerc *et al.*, 2003; Tharme, 2003; Annear *et al.*, 2004; Poff and Zimmerman, 2010; Mocq *et al.*, 2013). Hence, HSC remain key components of habitat models (Ahmadi-Nedushan *et al.*, 2006; Yi *et al.*, 2017).

HSC are built using physical and biological measurements from either a single river, hereafter referred to as *river-specific* HSC (Leclerc *et al.*, 1996; Guay *et al.*, 2000), or by combining data originating from multiple rivers, hereafter referred to as *regional* HSC (Lamouroux *et al.*, 1999; Mäki-Petäys *et al.*, 2002; Hedger *et al.*, 2004). While best practices for habitat modelling recommend the use of river-specific HSC, such HSC are not always available due to their costly and time-consuming development. Practitioners therefore often use HSC from other rivers or curves that were built using data and knowledge gathered at a regional scale. According to the literature, river-specific HSC have shown poor transferability to other rivers (Groshens and Orth, 1993; Leftwich *et al.*, 1997; Guay *et al.*, 2003; Strakosh *et al.*, 2003; Moir *et al.*, 2005), except for a few situations (Freeman *et al.*, 1997; Lamouroux *et al.*, 1999; Mäki-Petäys *et al.*, 2002). In some studies, river-specific HSC even had weak predictive power for different reaches of the same river on which they were developed (Scott and Shirvell, 1987; Bourgeois *et al.*, 1996; Guay *et al.*, 2000). In general, regional HSC increase the transferability between rivers but lead to relatively low predictive power (Hedger *et al.*, 2004). Given the poor performance and transferability of river or regional HSC, some studies have proposed the development of HSC from experts' knowledge rather than physical and biological measurements (Jorde *et al.*, 2001; Ahmadi-Nedushan *et al.*, 2008; Mocq *et al.*, 2013; Beaupré *et al.*, 2020). Expert models have

---

* Corresponding author at: INRS Eau Terre Environnement, 490 de la couronne, Quebec, PQ G1K 9A9, Canada.
E-mail address: jeremie.boudreault@inrs.ca (J. Boudreault).

https://doi.org/10.1016/j.ecolmodel.2022.109905
Received 21 September 2021; Received in revised form 27 January 2022; Accepted 8 February 2022
0304-3800/© 2021

shown good potential for model improvement when applied to new rivers (e.g. Mocq *et al.*, 2013). However, these models were mostly applied on only one or few rivers, which limits the conclusions that can be drawn in terms of performance and transferability.

HSC are most often developed from averaging habitat conditions across many sites in a stream (e.g. Guay *et al.*, 2000; Hedger *et al.*, 2004) or from experts' knowledge of "general" habitat preferences (e.g. Jorde *et al.*, 2001; Mocq *et al.*, 2013). Consequently, these models are not designed to account for site-specific differences that exist in available habitat that can cause local divergence in habitat selection (Rosenfeld *et al.*, 2011). When a stream or a reach has more of less suitable habitat because of the geomorphic conditions and/or fragmented habitat, fish may be forced to occupy suboptimal habitat at higher frequencies. This will result in a shift of the locally calibrated HSC toward a different optimum than those established in streams with a wider range of suitable habitat conditions, leading to a poor fit of HSC when transferred among sites or streams that differ in available habitat. Not considering local differences in available habitat may thus explain the low predictive power (e.g. Guay *et al.*, 2000; Hedger *et al.*, 2004) and relatively low transferability (e.g. Leftwich *et al.*, 1997; Guay *et al.*, 2003) of traditional habitat models. To our knowledge, all HSC that were considered in the literature have ignored these local differences in available habitat conditions. One justification for this oversight may be linked to the fact that systematically adapting models to various scenarios of available habitat conditions could become difficult, if not impossible, to apply in an operational context (i.e. in habitat model implementations).

Fortunately, recent developments in some statistical methods and their implementation in modern software (e.g. Ramsay and Silverman, 2005) open the possibility of producing site-specific HSC that naturally account for differences that exist in habitat conditions at smaller scales without complicating subsequent analyses. The actual definition of an HSC (i.e. a *curve*) makes it conducive to the development of a new approach using the functional data analysis (FDA) framework (Ramsay, 1982). FDA allows observations to be treated as curves or functions rather than scalar or vector observations in traditional contexts (Ramsay and Silverman, 2005). FDA has become very popular in recent years, both in hydrology (e.g. Chebana *et al.*, 2012; Masselot *et al.*, 2016; Ternynck *et al.*, 2016; Larabi *et al.*, 2017; Boudreault *et al.*, 2019) and ecology (e.g. Ainsworth *et al.*, 2011; Embling *et al.*, 2012; Stewart-Koster *et al.*, 2014; Yen *et al.*, 2015; Di Battista *et al.*, 2016; Boudreault *et al.*, 2021), enabling several types of novel analyses such as functional clustering (Jacques and Preda, 2014), functional analysis of variance (Górecki and Smaga, 2019) and functional regression (Morris, 2015).

Functional regression models (FRM) appear to be the most appropriate tool in our context by allowing to model a response variable *curve* (like an HSC) from inputs that are either scalar values (scalar-on-function models), curves (function-on-function models), or both (Morris, 2015). FRM can thus capture the variation in HSC between reaches by creating a function that modifies the shape of the HSC (the response in FRM) based on reach-specific predictors (e.g. depth, velocity) related to intersite variation in available habitat. This makes FRM ideally suited for modelling HSC based on available habitat conditions.

Furthermore, function-on-function FRM allows the predictor variable to also be included in the model as a *curve*. Hence, it becomes possible to create a model that links the HSC (response *curve*) to the local distribution of available habitat conditions (predictor *curve*). By design, this FRM accounts for local differences that exist in available habitat conditions compared to traditional HSC construction that aggregates habitat conditions and preferences across all sampled sites. Indeed, traditional model fitting generate a single average curve to predict HSC using data integrated across multiple sites that most likely differ in available habitats and their use. On the other hand, FRM can fit *individual* curves to each site, allowing for local prediction of HSC that functionally accounts for differences in available habitat, thus addressing an important challenge in fish habitat modelling. The benefits associated with site-specific HSC could lead to better performance at both river and regional scales (i.e. increased performance and transferability) and less field work required to develop locally adapted HSC based on available habitat.

In this study, we apply functional regression to build predictive models for site-specific HSC using juvenile Atlantic salmon (*Salmo salar*) as an example. To test and compare the proposed approach, traditional HSC are also developed at river and regional scales.

2. Material and methods

2.1. *Data collection*

Data on juvenile Atlantic salmon (*Salmo salar*) were collected during the summer of 2017 on two rivers located in the province of Quebec, Canada (Figure 1). The first river is the Sainte-Marguerite River, which flows into the Saguenay River, a tributary of the St-Lawrence River. The Sainte-Marguerite River consists of two main branches (Main and Northeast) of similar sub-basins sizes and a total watershed of 2118 km^2. Discharge was only recorded on the Northeast Branch (48°16′5" N, 69°54′33"W), ~1 km from the Main Branch confluence with the Northeast Branch and ~3 km from the Sainte-Marguerite's mouth, with a mean annual value of 30.1 m^3/s (based on records from 1999 to 2019). All measurements in this study were made on the Main Branch. The second river is the Petite-Cascapedia River. It flows into the Baie des Chaleurs in the Gaspésie region. At its gauging station (48°13′56" N, 65°43′59 W), located ~11 km upstream of its mouth, it has a drainage basin area of 1340 km^2 and a mean annual discharge of 30.6 m^3/s (based on records from 1982-2004 and 2012-2020). For each river, several small reaches (hereafter referred to as *sites*) were surveyed with biological and physical measurements as described in section 2.1 of Boudreault *et al.* (2021) for the Sainte-Marguerite River and section 2.4.2 of Beaupré *et al.* (2020) for the Sainte-Marguerite and Petite-Cascapedia rivers. Each site consisted of 30 habitat patches of 4 m^2 equally distributed along 5 transects i.e. six patches per transect (Figure 1). Studied sites were ~75 m long and had the same width as the river. All sites were at least 500 m apart along the river to ensure independence. Sites were semi-systematically selected to cover a wide diversity of habitat conditions such as pools, riffles, glides, etc. Habitat patches were fished using a Smith-Root LR24 electrofisher with a single-pass open net technique (Beaupré *et al.*, 2020; Boudreault *et al.*, 2021). All fish caught were identified to species (except for cyprinids), but only juvenile salmon were measured at fork length and then classified as fry (0+) or parr (1+, 2+) prior to release. On the Sainte-Marguerite River, juvenile salmon 30–54 mm-long were classified as 0+, 55–87 mm-long as 1+ and 88–125 mm-long as 2+ (Figure 2a). On the Petite-Cascapedia River, fish 42–62 mm-long were classified as 0+, 63–92 mm-long as 1+ and 93–132 mm-long as 2+ (Figure 2b). For all patches, physical measurements such as water depth, mean flow velocity and median substrate size (D50) were measured after electrofishing.

To ensure that enough fish were observed to build site-specific HSC, only sites with at least four observed juvenile salmon were kept for the analysis. This criterion was chosen as a trade-off between 1) having as many sites as possible and 2) having enough observations at each site to build credible site-specific HSC. This criterion has led us to combine data for parr of ages 1+ and 2+ (hereafter referred to as *parr*). Indeed, parr was the juvenile Atlantic salmon life stage having the highest number of sites fulfilling our criterion with $n = 36$ sites compared to fry ($n = 30$), 1+ parr ($n = 28$) and 2+ parr ($n = 8$). Hence, this study solely focused on parr and considered 1+ and 2+ parr together, as done in recent literature (e.g. Mocq *et al.*, 2018; Beaupré *et al.*, 2020).

To develop the models presented in the following sections, the three most important and commonly studied habitat variables for juvenile Atlantic salmon habitat were considered: depth, D50 and velocity (e.g. Guay *et al.*, 2000; Hedger *et al.*, 2004; Mocq *et al.*, 2013). These mea-

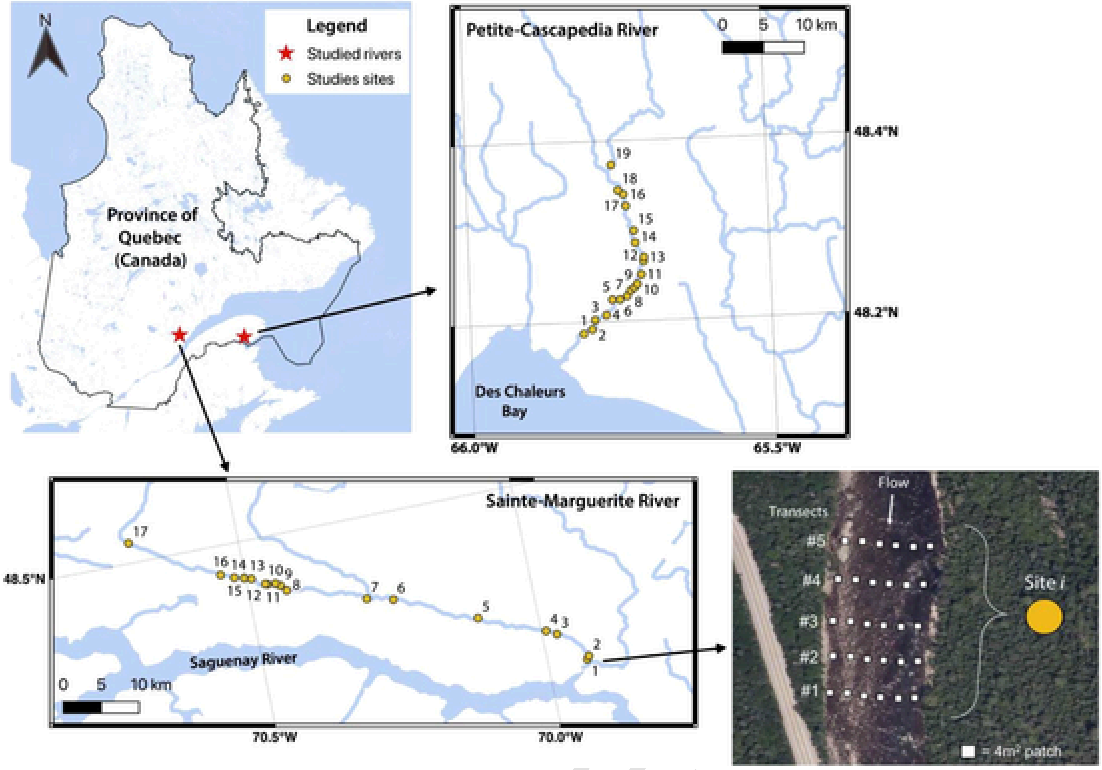


Fig. 1. Location of the studied rivers and the sampling design at each site.

sured physical habitat variables were then categorized into two classes: *available* and *selected*. The first class (*available*) corresponded to all physical measurements, regardless of parr presence or absence. The second class (*selected*) corresponded only to the physical measurements at patches where parr were observed. Summary statistics of physical and biological measurements collected on the two rivers are reported in Table 1.

2.2. Traditional habitat suitability curves (HSC)

HSC are defined in different ways in the literature, ranging from experts' knowledge (category I), frequency of habitat use (category II) and preference (category III) (Bovee *et al.*, 1998). Preference HSC are favored against frequency of use HSC because they are corrected for available habitat characteristics (Manly *et al.*, 2007). Various methods exist to derive preference HSC such as forage ratio (Bovee *et al.*, 1998), Jacobs index (Jacobs, 1974) and Ivlev electivity index (Ivlev, 1961). Forage ratio is the most popular and familiar index of preference according to Bovee *et al.* (1998) and Jowett *et al.* (2008). It is calculated by dividing the frequency of selected habitat by the frequency of available habitat variable in each class of a variable and by normalising the result to a maximum value of 1 that represents the most preferred habitat condition (Manly *et al.*, 2007). Forage ratios were used in this study (hereafter referred to simply as HSC).

Frequencies of selected and available habitat were estimated by kernel density estimates (KDE) (Hayes and Jowett, 1994; Jowett *et al.*, 2008). A KDE is defined as follows (Tukey, 1977):

$$f(s) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{s - s_i}{h}\right) \tag{1}$$

where $f(s)$ is the estimated KDE, $s_1, \ldots, s_n$ are univariate independent and identically distributed random variables with unknown probability density function, n is the number of observations, K is the kernel (often the standard normal density function) and h is called the bandwidth and is a parameter that can be adjusted for smoothness (Rudemo, 1982). Before adjusting KDE, habitat variables were right-censored to 99% percentile of their whole distribution to avoid right wavy tails in the KDE estimates, where only very few observations were made. For consistency of the KDE with the observed empirical distributions, KDE were plotted along with the histograms to validate the choice of the bandwidth parameter as done in Boudreault *et al.* (2021). A unique bandwidth value was chosen for each habitat variable. HSC were computed by dividing the frequency of selected habitat by the frequency of available habitat and then scaled to a maximum value of 1 as explained above (Hayes and Jowett, 1994).

For each of the considered habitat variable (depth, D50 and velocity), three HSC were developed based on measurements from 1) the Sainte-Marguerite River (HSC-SM), 2) the Petite-Cascapedia River (HSC-PC) and 3) both rivers (HSC-REG). The objective behind developing HSC at various scales (i.e. river and regional) was to correctly assess performance of traditional HSC and their transferability to other rivers compared to the new introduced method described below.

Finally, once univariate HSC are built, they can be combined into a unique multivariate habitat suitability index (HSI) using multiplication (e.g. Vadas and Orth, 2001), arithmetic or geometric means (e.g. Gard, 2014) or other more sophisticated ways such as regression (e.g. Guay *et al.*, 2000). HSI were computed for each patch using the standard method of multiplication to further validate our results.

2.3. Functional data analysis (FDA) and functional regression models (FRM)

In FDA, curves or functions represent the response variable to be modelled and/or the predictor variable(s) and need to be constructed from the observed discrete data. A usual initial step in FDA is to fit a known function (e.g. a linear combination of basis functions) to the observations (Ramsay and Silverman, 2005):

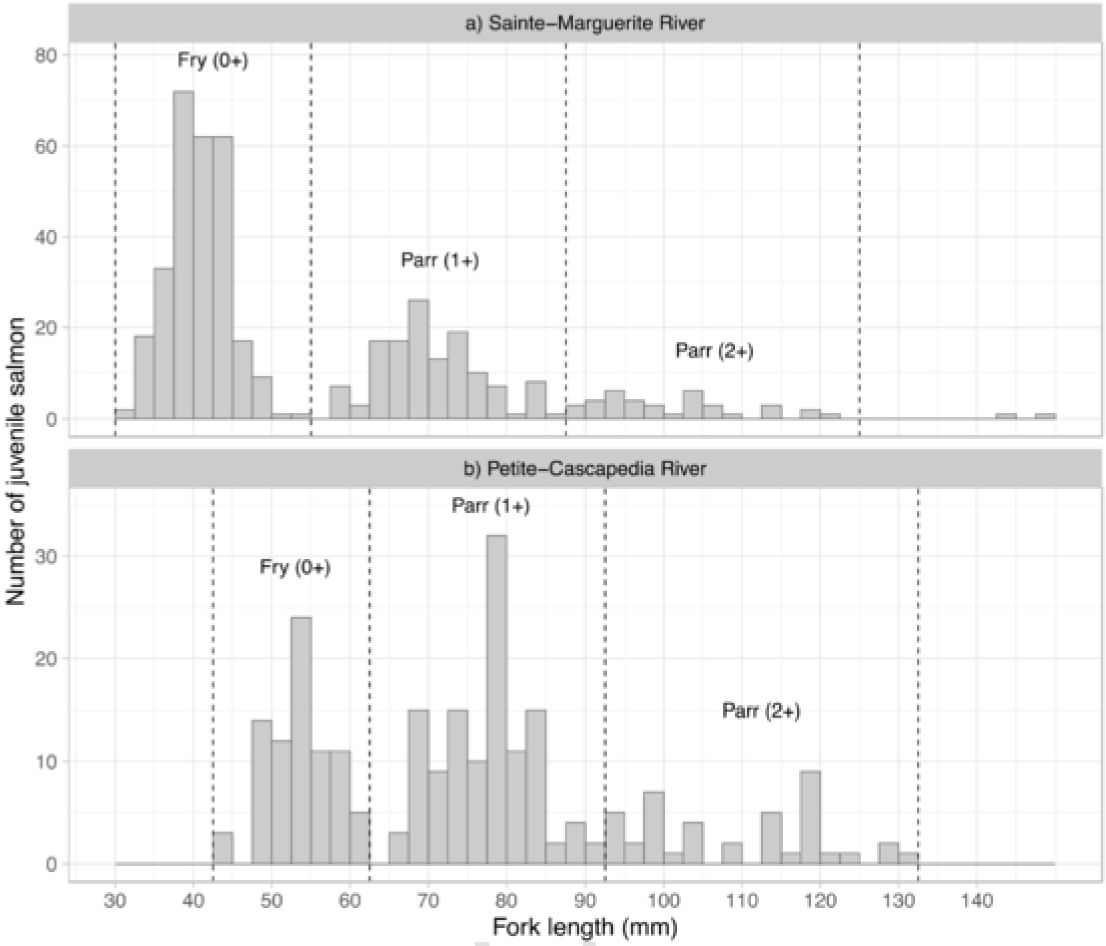


Fig. 2. Fork lengths distribution of juvenile Atlantic salmon.

Table 1
Overview of the studied rivers and sites.

River	Survey period	Number of sites	Mean physical measurements						Biological measurements	
			Available			Selected				
			Depth (cm)	D50 (mm)	Velocity (m/s)	Depth (cm)	D50 (mm)	Velocity (m/s)	Mean # of parr per site	Total # of parr
Sainte-Marguerite River	06/07/2017 to 16/08/2017	17	40 (±16)	95 (±43)	0.36 (±0.23)	40 (±17)	96 (±38)	0.31 (±0.19)	26.6 (±14.5)	453
Petite-Cascapedia River	31/08/2017 to 17/09/2017	19	38 (±19)	54 (±16)	0.47 (±0.32)	38 (±18)	57 (±14)	0.39 (±0.27)	22.7 (±13.1)	432
Combined	-	36	39 (±18)	73 (±38)	0.42 (±0.28)	39 (±18)	77 (±33)	0.35 (±0.23)	24.6 (±13.7)	885

$$f(s) = \sum_{m=1}^{M} c_m \varphi_m(s), \ s \in \Omega \tag{2}$$

where $f(.)$ is a continuous function, $\phi_m(.)$ are the known basis functions, c_m are the coefficients to be estimated, M is the number of basis functions and Ω is the domain over which $f(.)$ is constructed. Equation (2) is the traditional way to estimate a functional observation from discrete observations. However, new approaches for fitting FRM can now directly use the observed points on the continuous function without the need to first estimate the underlying continuous function (Brockhaus *et al.*, 2015). Hence, this prior step was not performed.

For the current application of FRM, there were two curves to be estimated for each site *i:* the response curve $Y_i(.)$, which represents the local HSC at site *i*, and the predictor curve $X_i(.)$, which represents the local available habitat. A natural choice for $X_i(.)$ was to consider the KDE of the available habitat conditions for the same variable used to build the HSC. For example, the KDE of available depths was used as a predictor to model the HSC for depth. The logic here is that depths chosen by fish – and, therefore, the empirical HSC for depth – are influenced by the depths that are available, which becomes the predictor function. The same applies for the velocity and D50 models.

Function-on-function FRM were used (Morris, 2015). The simplest model of this class is called the concurrent model (e.g. Goldsmith and Schwartz, 2017). It is the equivalent of the linear model but with both response and predictor functional variables:

$$Y_i(s) = \alpha(s) + \beta(s) X_i(s) + \varepsilon_i(s), \ s \in \Omega \tag{3}$$

where $Y_i(.)$ and $X_i(.)$ represent respectively the HSC and the KDE of available habitat conditions for a given habitat variable for the studied site $i = 1, 2, \ldots, 36$, Ω is the domain over which $Y(.)$ and $X(.)$ are observed and $\varepsilon_i(.)$ is the error term. In this model, the intercept $\alpha(.)$ and the regression coefficient $\beta(.)$ are also functions. Their definition is the same as *f(.)* in Eq. (2), a linear combination of known basis functions and coefficients to be estimated. Model (3) only allows the predictor $X_i(.)$ at point *s* or at a fixed lag *s – l* to be used to predict $Y_i(.)$ at *s* (Ramsay and Silverman, 2005). In order to take into account the complete information on the available habitat conditions, all the values of the predictor curves need to be used to predict *Y*(.) at *s*. Hence, the fully functional linear model is introduced:

$$Y_i(s) = \alpha(s) + \int_{\Omega_2} \beta(r,s) X_i(r)\, dr + \varepsilon_i(s),\ s \in \Omega_1 \tag{4}$$

where $\beta(r, s)$ is now a surface that represents the effect of $X_i(.)$ at any point *r* on Y_i and Ω_1 and Ω_2 are respectively the domains of $Y_i(.)$ and $X_i(.)$ (in our case $\Omega_1 = \Omega_2$). In model (4), all the values of $X_i(.)$ are now being used to model the value of $Y_i(.)$ at *s*, which is more adapted to our current study than model (3). One can note that model (3) is a special case of model (4), where $\beta(r,s)$ only takes values on the diagonal $s = r$. Hence, only model (4) was considered in this study (hereafter referred to simply as FRM).

All analyses were conducted in R (R Core Team, 2021). To calibrate FRM defined in (4), the *FDboost* package was used (Brockhaus *et al.*, 2017b). The parameters of the regression coefficients were estimated with a component-wise gradient boosting algorithm and the minimized loss function was the mean square error (Bühlmann and Hothorn, 2007; Brockhaus *et al.*, 2017a):

$$\sum_{i=1}^{n} \int_{\Omega_1} \left[Y_i(s) - \hat{Y}_i(s) \right]^2 ds \tag{5}$$

Optimal parameters of the model were found with a method called early stopping (Brockhaus *et al.*, 2017a). Briefly, the iterative procedure was stopped when the mean square error calculated with cross-validation stops decreasing and begins to increase. Early stopping avoids overfitting the model and leads to more regularized coefficients (Brockhaus *et al.*, 2017c). To assess the uncertainty of the estimated effects, standard error of the intercept and the regression coefficients surface were computed using bootstrapping (with 1000 replications) as done in Masselot *et al.* (2018).

It can be seen that traditional HSC described in section 2.2 can be defined as a very simple FRM (such as equation 3) with no predictor and only an intercept function $\alpha(.)$. This similarity demonstrates that the FDA framework is well suited for HSC construction and that performance metrics described in the next section can be calculated for both FRM and traditional HSC.

2.4. Performance criteria

To assess the performance of the models to predict site-specific HSC, two functional equivalents of the traditional R^2 were calculated (Ramsay and Silverman, 2005). The first one, denoted fR^2, measures the percentage of variation in the data explained by the model, adapted for functional observations. fR^2 is calculated as follows:

$$fR^2 = 1 - \frac{\sum_{i=1}^{n} \int_{\Omega} \left(Y_i(s) - \hat{Y}_i(s) \right)^2 ds}{\sum_{i=1}^{n} \int_{\Omega} \left(Y_i(s) - \bar{Y} \right)^2 ds} \tag{6}$$

where $Y_i(s)$ and $\hat{Y}_i(s)$ are respectively the observed and predicted response curve, $\bar{Y}$ is the mean observation, Ω is the domain over which *Y*(.) is observed and *n* is the number of functional observations. In addition to the fR^2, an alternate version was also calculated ($altfR^2$). In case studies working with functional data, it is of interest to know how the model performs compared to a model containing only a *functional* intercept $\alpha(.)$. In our case, the model with the functional intercept only is the average HSC that does not account for variation among sites as explained above. $altfR^2$ was computed as follows:

$$altfR^2 = 1 - \frac{\sum_{i=1}^{n} \int_{\Omega} \left(Y_i(s) - \hat{Y}_i(s) \right)^2 ds}{\sum_{i=1}^{n} \int_{\Omega} \left(Y_i(s) - \bar{Y}(s) \right)^2 ds} \tag{7}$$

where the mean observation $\bar{Y}$ in the denominator of equation (6) is replaced the mean *functional* observation $\bar{Y}(s) = \sum_{i=1}^{n} Y_i(s)/n$. Note that in the functional context, both fR^2 and $altfR^2$ can take on negative values if the fit is worse with $\hat{Y}_i(s)$ compared to $\bar{Y}$ and $\bar{Y}(s)$, respectively (Ramsay and Silverman, 2005).

The two above criteria were also calculated with cross validation. Given that the number of studied sites was small ($n = 36$ functional observations), a leave-one-out cross-validation procedure was performed (Quenouille, 1949). Basically, one site was removed from the dataset prior to fitting the model. Then, the resulting model was used to predict the HSC at the removed site. This procedure was repeated for all sites. To calculate performance measures with leave-one-out cross validation, $\hat{Y}_i(s)$ was replaced by $\hat{Y}_{(-i)}(s)$ in Eqs. (6) and (7) where $\hat{Y}_{(-i)}(s)$ is the predicted HSC for site *i*, when this site was removed from the dataset prior to fitting the model.

Apart from cross-validation, all performance metrics were calculated separately for the sites studied on the Sainte-Marguerite River, the Petite-Cascapedia River and both simultaneously. These additional results were helpful to check the adequacy of the models on the same river as they were calibrated, but also on the river on which they were not developed. This refers to the concept of transferability as discussed in the introduction of the paper. As a final validation of the developed HSC, HSI values were regressed against the observed fish density (e.g. Guay *et al.*, 2000; 2003; Beaupré *et al.*, 2020). R^2 values were reported as a measure of the ability of HSI to explain fish density.

3. Results

3.1. Traditional habitat suitability curves (HSC) results

Traditional habitat suitability curves are reproduced in Fig. 3. On the Sainte-Marguerite and Petite-Cascapedia rivers, two preferred depths ranges were noted (Fig. 3, left panels): 20 – 25 cm and 55 – 65 cm. Shallow waters (< 15 cm) were mostly avoided by parr. These results were consistent across rivers. Preferred D50 values ranged from 60 mm to 130 mm on the Sainte-Marguerite River (Fig. 3, top-center). On the Petite-Cascapedia River, the available substrate sizes were smaller than on the Sainte-Marguerite River. However, the HSC of the Petite-Cascapedia River also showed that parr appeared to prefer larger (80 – 110 mm) substrate sizes (Fig. 3, middle-center). After combining measurements from the Sainte-Marguerite and Petite-Cascapedia rivers for the regional model, the 60 – 130 mm range remained the most preferred D50 values among the two rivers (Fig. 3, bottom-center). Parr mostly avoided smaller substrates (< 50 mm). For velocity, preferred values were between 0.25 and 0.35 m/s (Fig. 3, right panels) on both rivers. Velocities greater than 1 m/s were less available and did not seem to be preferred by parr.

3.2. Functional regression models (FRM) results

Resulting HSC and KDE for three sites (two on Petite-Cascapedia River and one on Sainte-Marguerite River) along with histograms of available and selected habitat characteristics are reproduced in Fig. 4.

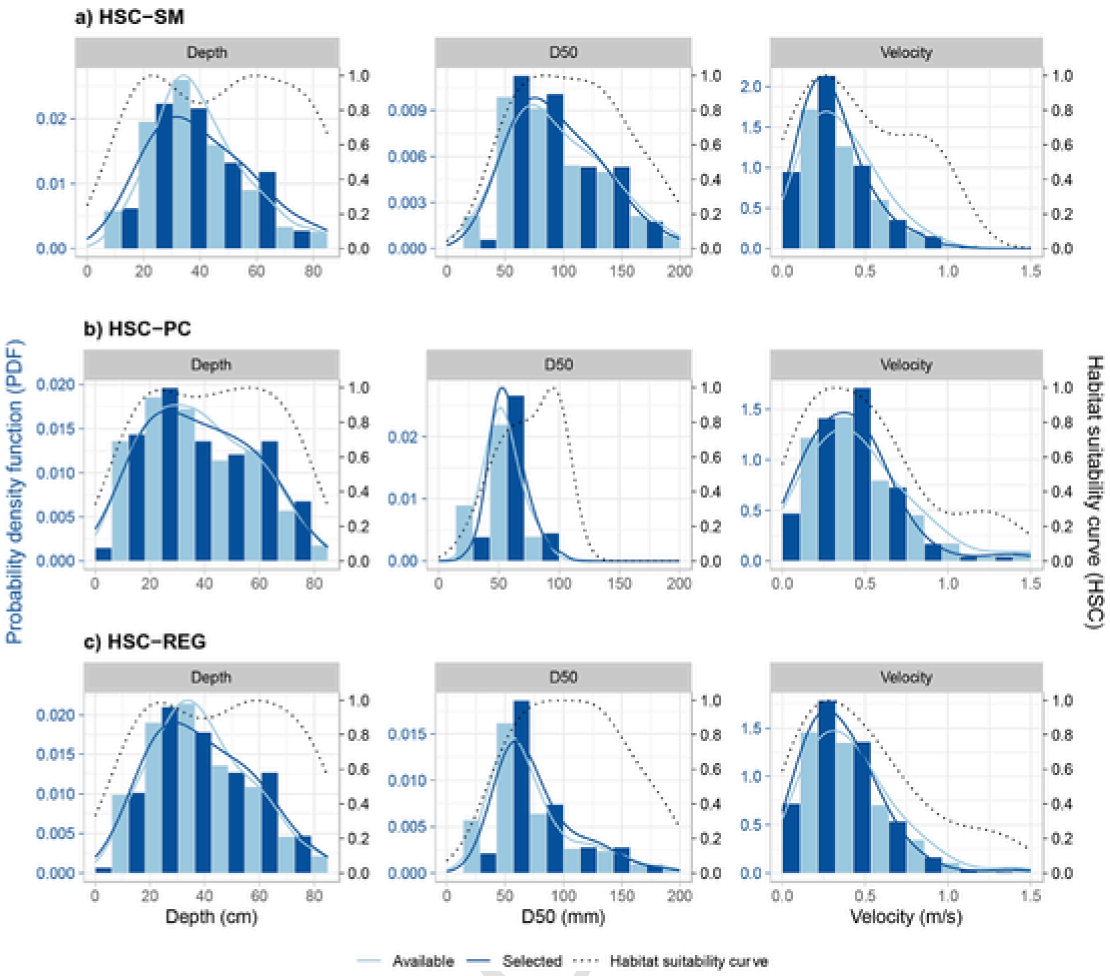


Fig. 3. Traditional habitat suitability curves (HSC) for depth, median substrate size (D50) and velocity developed on the a) Sainte-Marguerite River (HSC-SM), b) the Petite-Cascapedia River (HSC-PM) and c) on both rivers (HSC-REG). The light blue color represents the available habitat, the dark blue color represents the selected habitat, and the dotted black line represents the HSC.

An illustration of all HSC and KDE is available in supplementary material. KDE of available habitat and HSC were used as direct inputs for FRM described in this section.

FRM for the three habitat variables are reproduced in Fig. 5 for the intercept (left panels), the regression coefficients surface (center panels) and its standard error (right panels). To correctly read and interpret the FRM regression surfaces and their standard errors, some explanations must first be given. The *x*-axis corresponds to the values of the available habitat variable, while the *y*-axis corresponds to the values of the HSC. The regression coefficient surface shows the relationship between an increase or a decrease in the availability of the habitat variable (*x*-axis) and the HSC for values on the *y*-axis. A red region indicates a positive relation between availability and HSC: an increase (decrease) in habitat variable availability is correlated to an increase (decrease) in the HSC. On the other hand, a blue region indicates a negative relation. The conditional tense is used through the results as these relations are correlation rather than causation.

3.2.1. *FRM for depth*

As for traditional HSC, FRM for depth also showed that two ranges of depths were preferred by parr (25 – 35 mm and 50 – 60 mm) and that both shallowest (< 10 cm) and deepest (> 70 cm) waters were avoided, as illustrated by the intercept function (Fig. 5, top-left). The red region at the bottom-left of the surface indicates that increasing availability of shallow water depths (0 – 15 cm) may be linked to an increased HSC for slightly higher values of depths (0 – 25 cm) (Fig. 5, top-center). Also, available depths in the range 60 – 80 cm seemed to favor higher values of HSC for depths in the range 40 – 80 mm, while reducing the HSC for shallower depths (0 – 20 cm). The standard error of the surface was relatively low (Fig. 5, top-right).

3.2.2. *FRM for median substrate size (D50)*

According to the intercept function, the range of preferred D50 values for parr was from 50 mm to 70 mm (Fig. 5, middle-left). Parr seemed to avoid low (< 25 mm) and high (> 175 mm) values of D50. The regression coefficients surface showed that availability of substrate ranging from 100 mm to 200 mm tended to increase the HSC for such values and decrease the HSC for smaller substrate sizes (< 100 mm, blue region on the surface) (Fig. 5, middle-center). A smaller effect was the increased availability of small substrates (10 – 60 mm) that tended to reduce the HSC for higher values (70 mm to 150 mm). The standard error of the surface was greater for high values of available substrate sizes (i.e. > 170 mm) (Fig. 5, middle-right).

3.2.3. *FRM for velocity*

Velocities from 0.20 m/s to 0.40 m/s were preferred by parr as shown by the intercept function (Fig. 5, bottom-left). Faster currents (> 1 m/s) were mostly avoided. The regression coefficient surface indicated that increasing the availability of high velocities (> 1 m/s) may increase the HSC for slightly smaller values of velocities in the range of 0.40 m/s to 1 m/s as illustrated by the red region (Fig. 5, bottom-center). The regression coefficient surface was close to 0 else-

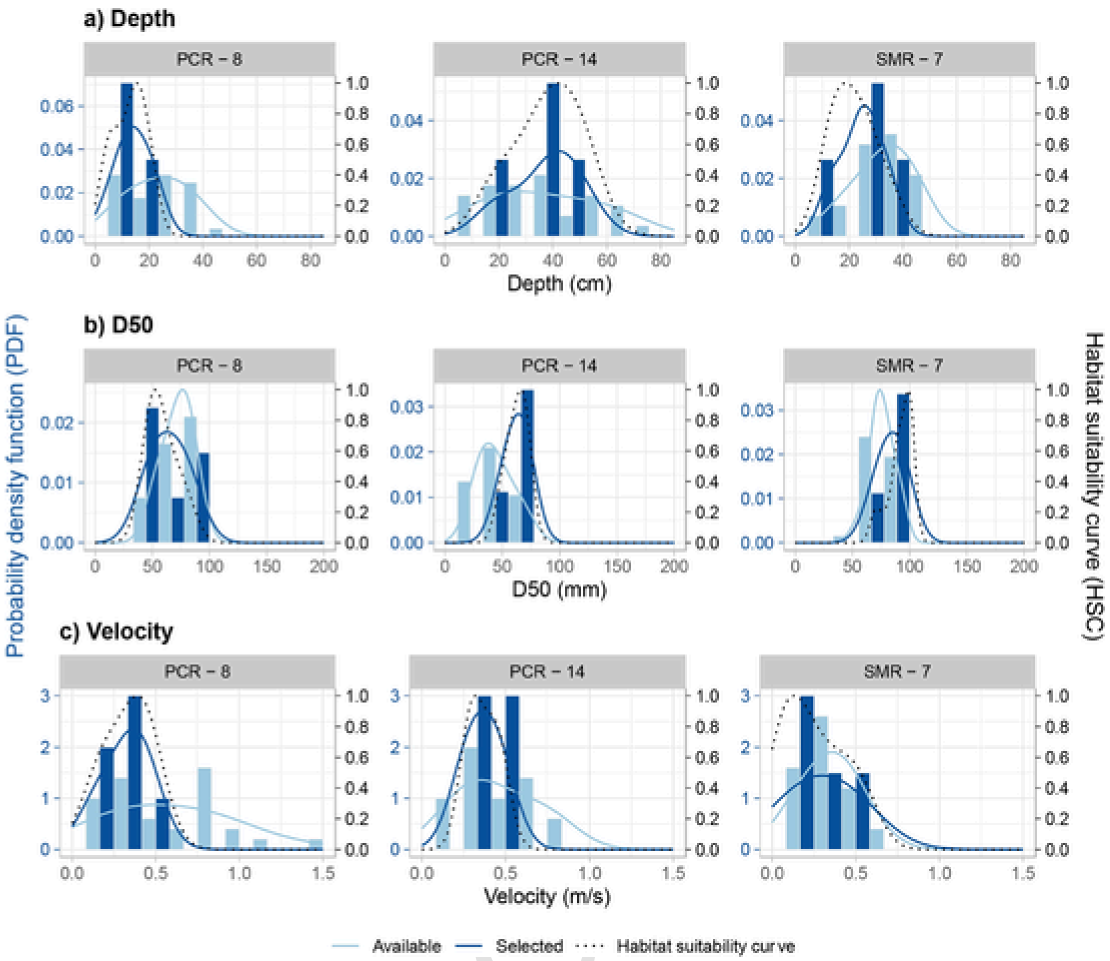


Fig. 4. Habitat suitability curve (HSC) and kernel density estimates (KDE) at three studied sites for a) depth, b) median substrate size (D50) and c) velocity. The light blue color represents the available habitat, the dark blue color represents the selected habitat, and the dotted black line represents the HSC.

where. For very high values of velocities (> 1.25 m/s), the regression coefficient surface had a higher standard error (Fig. 5, bottom-right).

3.3. Performance metrics and comparison

3.3.1. Site-specific habitat suitability curves (HSC) performance

The performance of site-specific HSC modelling expressed by fR^2 values are reported in Table 2. The performance on the Sainte-Marguerite River was the highest with the FRM for all three habitat variables (fR^2 of 0.480, 0.543 and 0.511 respectively for depth, D50 and velocity). For all traditional HSC (HSC-SM, HSC-PC and HSC-REG), these models performed poorly to explain variation in site-specific HSC with negative values of fR^2. These results suggest that traditional river or regional HSC are not adapted for local differences in site-specific HSC. On the Petite-Cascapedia River, the same conclusions apply: FRM explained a large proportion of the variability for all three site-specific HSC (fR^2 of 0.278, 0.499 and 0.404 for depth, D50 and velocity), while all traditional HSC performed poorly. Among traditional HSC applied to the Petite-Cascapedia River, the HSC developed on the Petite-Cascapedia River (HSC-PC) performed better than the others for depth and D50, while the HSC developed regionally (HSC-REG) performed better for velocity. However, all fR^2 values were negative for traditional HSC, so HSC developed at the river or regional scale did not fit well site-specific HSC on the Petite-Cascapedia River. When all sites from the two rivers were considered, the FRM again performed better than regional HSC at modelling site-specific HSC, with higher fR^2 values (0.380, 0.533 and 0.455 for depth, D50 and velocity) versus negative values for the regional HSC. When looking at the performance metrics calculated with the leave-one-out cross validation, the same conclusions apply, suggesting that FRM were not overfitted. In conclusion, FRM correctly modelled the differences in local HSC observed at each site (with high values of fR^2), while none of the traditional HSC was able to represent these local differences in site-specific HSC.

To further validate the different models, $altfR^2$ was also computed (Table 2), with lower values than fR^2 values. Recall that $altfR^2$ compares the predicted HSC to a mean functional observation, which already explains a great proportion of the variability in HSC. When comparing traditional HSC and FRM based on $altfR^2$ values, the same conclusions as above apply. Regarding FRM, $altfR^2$ highlighted the added value of considering available habitat in the models versus using only a functional intercept with $altfR^2$ values ranging from 0.036 to 0.414. Both depth and D50 FRM obtained higher values of $altfR^2$ when applied to one or two rivers. However, two FRM had lower $altfR^2$ values: the D50 FRM applied to the Petite-Cascapedia River ($altfR^2$ of 0.036) and the velocity FRM applied to the Sainte-Marguerite River ($altfR^2$ of 0.026). Such results may signify that local differences in available substrate and velocity values among sites were lower on the Sainte-Marguerite and the Petite-Cascapedia rivers, respectively, such that parr were exposed to the optimal values that they could use at will. In cross validation, the $altfR^2$ for the velocity FRM was slightly negative, meaning that the velocity FRM may not be better than an FRM with a unique intercept function. This was also the case for the D50 FRM, but only when applied to the Petite-Cascapedia River. This FRM was still relevant on the sites of the Sainte-Marguerite River, for which the value of $altfR^2$ in cross validation was higher (0.283).

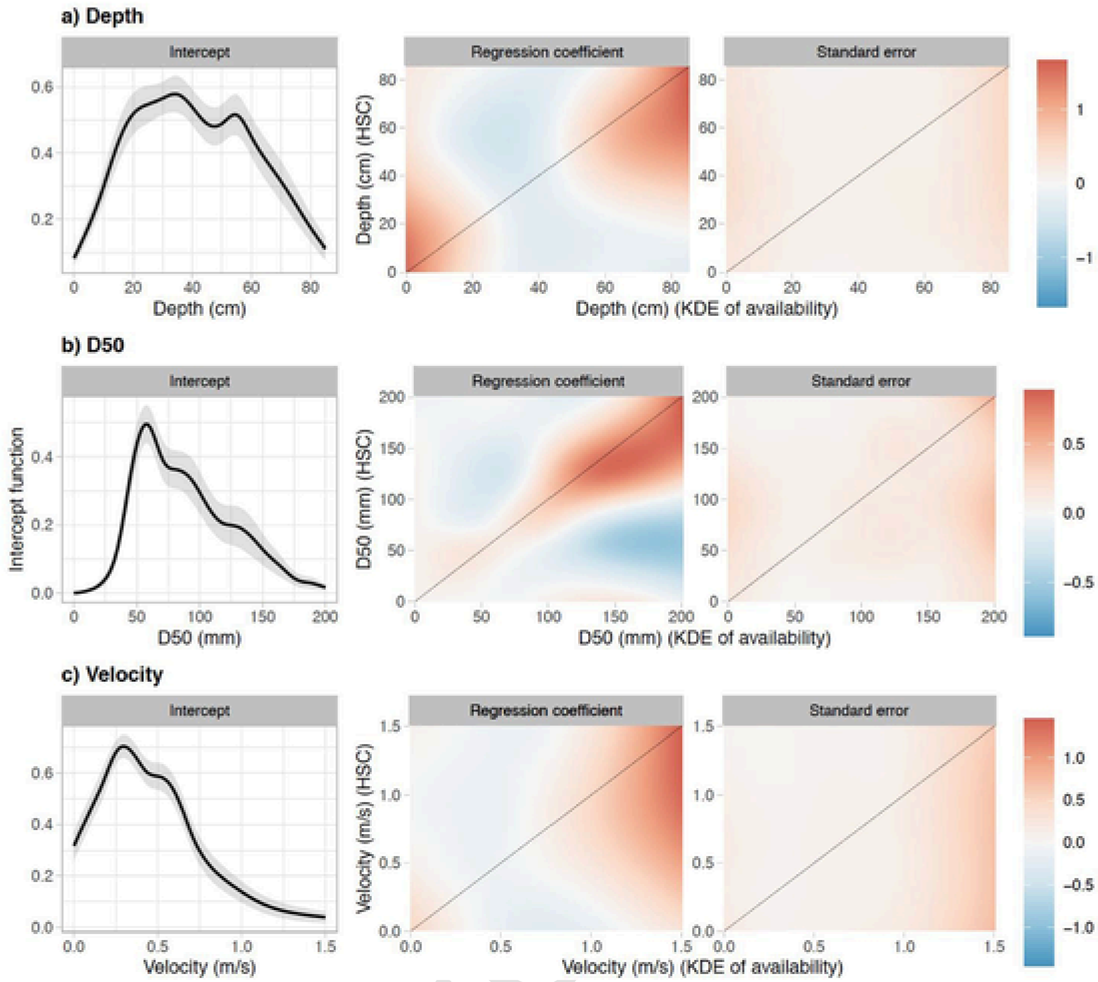


Fig. 5. Functional regression models (FRM) coefficients for a) depth, b) median substrate size (D50) and c) velocity. The intercept function α(s) (left panels) shows the general trend in HSC and its standard error is illustrated in light gray. The regression coefficients surface β(r, s) (center panels) show how an increase (or a decrease) in the availability of the habitat variable is correlated to an increase (in red) or decrease (in blue) in the HSC for a given value of the habitat variable. The standard error of the regression coefficients (right panels) shows the region of high variability in the estimates of the regression surfaces.

3.3.2. Habitat suitability index (HSI) performance

HSI values were computed at each surveyed patch of 4 m^2 and regressed against observed parr density. R^2 values for this relationship are reported in Table 3. While all R^2 values were very low (i.e. below 0.05), traditional HSC seems to perform slightly better than FRM to explain local fish density at each habitat patch when aggregated into HSI. On the Sainte-Marguerite River, the R^2 between HSI and parr density was 0.029, which is slightly higher than the results of the regional HSC (R^2 of 0.022) and of the FRM (R^2 of 0.018). On the Petite-Cascapedia River, the R^2 between HSI and fish density was 0.040, while it was 0.037 for the regional HSC and 0.030 for the FRM. It can be noted that on each river separately, river-specific HSC performed better than HSC developed on the other river in explaining fish density. Both traditional HSC and FRM regional models performed better, but did not outperform river-specific HSC to explain parr density. When both rivers were considered, R^2 was slightly higher for the traditional HSC than for the FRM.

4. Discussion

Our results showed that functional regression models (FRM) were powerful tools for constructing site-specific habitat suitability curves (HSC) in small river reaches. FRM were used to model local HSC based on the distribution of available habitat conditions for three key habitat variables of Atlantic salmon parr: water depth, median substrate size (D50) and mean flow velocity. FRM explained a large proportion of the variability in site-specific HSC for parr (fR^2 between 38% and 53%). In comparison, FRM had greater performance to model site-specific HSC than traditional HSC developed at the scale of each river or regionally, which had negative fR^2 values. When looking at $altfR^2$ values (ranging from 10.7% to 39.5% when calculated on both rivers), the benefits of FRM were further demonstrated over traditional approaches. A cross validation (leave-one-out) showed that FRM still outperformed traditional HSC models. These results highlight that FRM have correctly represented the fact that different available habitat conditions within a river reach can lead to different habitat choices (therefore, different local HSC) and show the need to consider these divergences in HSC construction. This was only made possible by taking a new look at HSC through functional data analysis (FDA) and linking HSC to the distribution of available habitat conditions in FRM.

When looking at transferability of traditional HSC, performance on the Petite-Cascapedia River was generally better for the HSC developed with the measurements from the same river compared to the HSC developed with the measurements from the other river or the regional HSC. Indeed, these results were expected (e.g. Hedger *et al.*, 2004) as it is recommended to use river-specific HSC (when available) in habitat modelling. However, such results were not found on the Sainte-Marguerite River, where HSC from the Petite-Cascapedia River or regional HSC were more adapted to site-specific HSC than the HSC developed on the Sainte-Marguerite River. In all cases, the newly introduced FRM outper-

Table 2
Performance of all models.

		fR^2					
Models	**Tested river(s)**	**Training**			**Cross validation**		
		Depth	**D50**	**Velocity**	**Depth**	**D50**	**Velocity**
HSC-SM	Sainte-Marguerite River	-1.787	-2.268	-0.470	-1.715	-2.260	-0.553
HSC-PC		-1.576	-0.874	-0.382			
HSC-REG		-1.771	-2.351	-0.285	-1.660	-5.164	-0.237
FRM		**0.480**	**0.543**	**0.511**	**0.390**	**0.441**	**0.460**
HSC-PC	Petite-Cascapedia River	-1.463	-1.079	-0.226	-1.490	-1.133	-0.303
HSC-SM		-1.652	-4.976	-0.345			
HSC-REG		-1.634	-5.171	-0.175	-1.660	-5.164	-0.237
FRM		**0.278**	**0.499**	**0.404**	**0.193**	**0.446**	**0.320**
HSC-REG	Both	-1.692	-3.502	-0.216	-1.712	-3.497	-0.260
FRM		**0.380**	**0.533**	**0.455**	**0.292**	**0.454**	**0.386**
		altfR^2					
Models	**Tested river(s)**	**Training**			**Cross validation**		
		Depth	**D50**	**Velocity**	Depth	**D50**	**Velocity**
HSC-SM	Sainte-Marguerite River	-2.617	-3.192	-1.927	-2.524	-3.182	-2.093
HSC-PC		-2.344	-1.404	-1.753			
HSC-REG		-2.596	-3.298	-1.560	-2.117	-10.86	-0.856
FRM		**0.326**	**0.414**	**0.026**	**0.209**	**0.283**	**-0.076**
HSC-PC	Petite-Cascapedia River	-1.886	-3.001	-0.839	-1.918	-3.103	-0.954
HSC-SM		-2.107	-10.498	-1.017			
HSC-REG		-2.087	-10.874	-0.762	-2.117	-10.86	-0.856
FRM		**0.153**	**0.036**	**0.106**	**0.054**	**-0.066**	**-0.02**
HSC-REG	Both	-2.261	-4.822	-0.993	-2.285	-4.815	-1.066
FRM		**0.248**	**0.395**	**0.107**	**0.143**	**0.294**	**-0.007**

Performance of site-specific HSC modelling in terms of fR^2 and altfR^2 for depth, median substrate size (D50) and velocity for all modelling techniques: traditional habitat suitability curves developed on the Sainte-Marguerite River (HSC-SM), on the Petite-Cascapedia River (HSC-PC), on both rivers (HSC-REG) and functional regression models (FRM). Results are split by river and by type of validation (training and cross validation).

Table 3
Relation between habitat suitability index and parr density.

Model	R^2		
	Sainte-Marguerite River	Petite-Cascapedia River	Both
HSC-SM	0.029	0.029	-
HSC-PC	0.003	0.040	-
HSC-REG	0.022	0.037	0.030
FRM	0.018	0.030	0.024

Performance of the habitat suitability index (HSI) to explain parr density in terms of R^2 at each surveyed patch of the Sainte-Marguerite River, the Petite Cascapedia River or both using the traditional habitat suitability curves developed on the Sainte-Marguerite River (HSC-SM), on the Petite-Cascapedia River (HSC-PC), on both rivers (HSC-REG) and functional regression models (FRM).

formed all traditional HSC when applied to Sainte-Marguerite River, Petite-Cascapedia River or both. The fact that FRM, developed with the measurements of the two rivers combined, outperformed river-specific HSC is a promising finding in terms of improved transferability. Considering local differences in available habitat among sites or rivers in FRM may thus increase transferability of HSC, which is reported to be low for traditional HSC (e.g. Groshens and Orth, 1993; Leftwich *et al.*, 1997; Guay *et al.*, 2003; Strakosh *et al.*, 2003; Hedger *et al.*, 2004; Moir *et al.*, 2005). Despite better results for performance measured on site-specific HSC directly, the combined habitat suitability index (HSI) was not a good indicator of parr density ($R^2 < 0.05$ for all models considered). In fact, R^2 values were a little higher for traditional HSI (using river-specific HSC) than for HSI calculated from predicted site-specific HSC by the FRM.

FRM have allowed assessing the relation between the available habitat conditions and the resulting HSC. For example, an increase in available deep waters (> 60 cm) seemed to increase the HSC towards water depths ranging from 60 cm to 80 cm, while reducing the HSC for smaller depths (0 – 20 cm). Also, an increase in bigger substrate availability (> 100 mm) may reduce the HSC values for smaller substrate sizes (20 – 80 mm). These findings were only made possible by introducing FRM to model site-specific HSC. By employing more information (*curves*), FRM also provides new results that could not be obtained from traditional models. However, these associations will need to be further validated with more data in order to derive functional relationships (i.e. causation). Nonetheless, this study shows the potential of FRM to provide ecologically valuable information in habitat modelling.

While this study demonstrates the advantages of using FRM to build site-specific HSC, some future avenues must also be discussed. First, even though site-specific HSC were better predicted by FRM compared to traditional river or regional HSC, this improved performance did not translate into stronger relations between HSI and fish density (i.e. all R^2 values were below 0.05). Improving the HSI is another topic of interest in the literature because traditional ways to summarize HSC into HSI (e.g. multiplication, as used in this study) have been criticized for a long time for its mixed performance (e.g. Guay *et al.*, 2000, 2003) and its assumption of independence between individual HSC (Railsback, 2016). Our study, which introduced the FDA framework for HSC construction, opens the door to refine the HSI computation by using FRM to directly predict the HSI from habitat available conditions and taking into account interactions, building of the recent work of Boudreault *et al.* (2021).

Second, due to the low number of studied sites ($n = 36$), our FRM only included one predictor curve and thus considered HSC in a univariate manner. The sample size is a limitation of FDA and suggests continuing measuring and collecting data with recent technologies such as airborne sensors, video analysis or high-quality satellite images (e.g. Smith *et al.*, 2005; Hedger *et al.*, 2006; Dugdale *et al.*, 2013). As habitat selection depends on other habitat variables that need to be jointly considered (Railsback, 2016), multivariate FRM should be further exploited to produce more biologically representative HSC based on several predictors.

Third, all of our models (traditional HSC and FRM) were based on observed frequency of use habitat, which can be influenced a myriad of factors. Our electrofishing sampling technique (i.e. single pass with open parcels) can bring some bias in the obtained results (Bain *et al.*, 1985), even though it has shown good correlation with the multiple pass technique (Bateman *et al.*, 2005). Fish sizes will affect habitat needs, with greater sizes associated with preference toward slightly deeper waters and bigger substrates (Heggenes *et al.*, 1999). Such differences may, however, not impact too much our results as considering 1+ and 2+ parr together is common in the literature (e.g. Mocq *et al.*, 2013; Beaupré *et al.*, 2020). Other factors such as fish density, competition, territoriality and predation can influence habitat selection, and consequently, the resulting HSC (e.g. Gabler and Amundsen, 1999; Dionne and Dodson, 2002). Also, stream temperature could be considered in future developments of FRM, especially when the temperature exceeds a threshold and forces fish to migrate toward thermal refuges (Beaupré *et al.*, 2020). In our database, the intra-site variability of stream temperature was low (Boudreault *et al.*, 2021), which is the reason for not developing individual HSC for stream temperature. Recent methods also suggest the use of bioenergetics-based HSC against preference HSC (e.g. Rosenfeld *et al.*, 2016; Naman *et al.*, 2020). Nevertheless, preference HSC are still being widely used and the proposed methodology can, at least, cope with some of the drawbacks of traditional HSC.

Finally, the developed FRM have not yet been applied in an operational context to calculate, for example, the weighted usable area (e.g. Lamouroux et al., 1998; Mocq et al., 2013). Therefore, it is suggested that hydraulic models be coupled with FRM developed in this study, which could better serve the needs of river managers and decision makers by providing a better set of HSC adapted to local differences in available habitat conditions among sites or reaches.

5. Conclusion

Our study showed that functional data analysis (FDA) is a suitable framework for developing habitat suitability curves (HSC) that are key components of various fish habitat models. Indeed, FDA naturally treated HSC as *curves* allowing various *functional* statistical analyses without complicating the subsequent steps. We used functional regression models (FRM) to build models producing site-specific HSC that took into account the differences in local available habitat. Our study opens the door to various innovations and performance improvements for the biological component of habitat models. Therefore, it is suggested that the potential of FDA is further exploited to consider HSC modelling in a multivariate manner, to refine the HSI calculation or to combine the developed FRM with hydraulic models.

CRediT authorship contribution statement

Jeremie Boudreault: Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Writing – original draft, review & editing. **Normand E Bergeron:** Conceptualization, Methodology, Writing – review & editing, Resources, Project administration. **Andre St-Hilaire:** Conceptualization, Methodology, Writing – review & editing, Resources. **Fateh Chebana:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

The authors wish to acknowledge the financial contribution of Natural Sciences and Engineering Research Council of Canada for the scholarship of the lead author, to the Fonds de Recherche du Québec - Nature et Technologies for the team research grant (grant #208647) and to the Canadian Foundation for Innovation (CFI) for the equipment funding. The authors would like to thank André Boivin, Marc-André Pouliot, Killian Dolais, Andrée-Sylvie Carbonneau, Michael Deetjens and Antoine Boudry for their help in the field work. The authors are grateful to the editors-in-chief, the associate editor Ursula Scharler, Jordan Rosenfeld and another anonymous reviewer for their constructive comments which greatly improved the quality of the paper. The authors have no conflict of interest to declare.

Data availability statement

Data and codes to calibrate all models used in the study are made freely available on the first author GitHub page (http://github.com/jeremieboudreault).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2022.109905.

References

Ahmadi-Nedushan, B, St-Hilaire, A, Berube, M, Ouarda, T, Robichaud, E, 2008. Instream flow determination using a multiple input fuzzy-based rule system: A case study. River Res. Appl. 24 (3), 279–292.

Ahmadi-Nedushan, B, St-Hilaire, A, Bérubé, M, Robichaud, É, Thiémonge, N, Bobée, B, 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Res. Appl. 22 (5), 503–523.

Ainsworth, L, Routledge, R, Cao, J, 2011. Functional data analysis in ecosystem research: the decline of Oweekeno Lake sockeye salmon and Wannock River flow. J. Agricult., Biol., Environ. Statis. 16 (2), 282–300.

Annear, T C, Aarrestad, P, Locke, A, Beecher, H A, Chisholm, I, Mayes, K B, Marshall, J, Coomer, C E, Hunt, Jand, Jacobson, R, 2004. Instream flows for riverine resource stewardship. Instream Flow Council, Cheyenne, Wyoming.

Bain, M B, Finn, J T, Booke, H E, 1985. A quantitative method for sampling riverine microhabitats by electrofishing. North Am. J. Fisheries Manag. 5 (3B), 489–493.

Bateman, D S, Gresswell, R E, Torgersen, C E, 2005. Evaluating single-pass catch as a tool for identifying spatial pattern in fish distribution. J. Freshwater Ecol. 20 (2), 335–345.

Beaupré, J, Boudreault, J, Bergeron, Nand, St-Hilaire, A, 2020. Inclusion of water temperature in a fuzzy logic Atlantic salmon (Salmo salar) parr habitat model. J. Therm. Biol 87, 102471.

Boudreault, J, Bergeron, N E, St-Hilaire, A, Chebana, F, 2019. Stream temperature modeling using functional regression models. JAWRA J. Am. Water Res. Ass. 55 (6), 1382–1400.

Boudreault, J, St-Hilaire, A, Chebana, F, Bergeron, N E, 2021. Modelling fish physico-thermal habitat selection using functional regression. Journal of Ecohydraulics 1–16.

Bourgeois, G, Cunjak, R A, Caissie, Dand, El-Jabi, N, 1996. A spatial and temporal evaluation of PHABSIM in relation to measured density of juvenile Atlantic salmon in a small stream. North Am. J. Fisheries Manag. 16 (1), 154–166.

Bovee KD (1982) Guide to stream habitat analysis using the instream flow incremental methodology. Available from the National Technical Information Service, *Springfield VA 22161 as PB 83-131052*. Report.

Bovee KD, Lamb BL, Bartholow JM, Stalnaker CB, Taylor J and Henriksen J (1998) Stream habitat analysis using the instream flow incremental methodology. (U.S. Geologicial Survey, Biological Resources Division Information and Technology).

Brockhaus, S, Melcher, M, Leisch, F, Greven, S, 2017a. Boosting flexible functional regression models with a high number of functional historical effects. Statistic. Comput. 27 (4), 913–926.

Brockhaus S, Ruegamer D, Hothorn T and Brockhaus MS (2017b) Package ‘FDboost’.

Brockhaus S, Rügamer D and Greven S (2017c) Boosting Functional Regression Models with FDboost. *arXiv preprint arXiv:1705.10662*.

Brockhaus, S, Scheipl, F, Hothorn, T, Greven, S, 2015. The functional linear array model. Statistic. Modell. 15 (3), 279–300.

Bühlmann, P, Hothorn, T, 2007. Boosting algorithms: Regularization, prediction and model fitting. Statistic. Sci. 477–505.

Chebana, F, Dabo-Niang, S, Ouarda, T B, 2012. Exploratory functional flood frequency analysis and outlier detection. Water Resour. Res. 48 (4), W04514.

DeGraaf, D, Bain, L, 1986. Habitat use by and preferences of juvenile Atlantic salmon in two Newfoundland rivers. Trans. Am. Fish. Soc. 115 (5), 671–681.

Di Battista, T, Fortuna, F, Maturo, F, 2016. Parametric functional analysis of variance for fish biodiversity assessment. J. Environ. Inf 28 (2), 101–109.

Dionne, Mand, Dodson, J J, 2002. Impact of exposure to a simulated predator (Mergus merganser) on the activity of juvenile Atlantic salmon (Salmo salar) in a natural environment. Can. J. Zool. 80 (11), 2006–2013.

Dugdale, S J, Bergeron, N E, St-Hilaire, A, 2013. Temporal variability of thermal refuges and water temperature patterns in an Atlantic salmon river. Remote Sens. Environ. 136, 358–373.

Embling, C B, Illian, J, Armstrong, E, van der Kooij, J, Sharples, J, Camphuysen, K C, Scott, B E, 2012. Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. J. Appl. Ecol. 49 (2), 481–492.

Freeman, M C, Bowen, Z H, Crance, J H, 1997. Transferability of habitat suitability criteria for fishes in warmwater streams. North Am. J. Fisheries Manag. 17 (1), 20–31.

Gabler, H M, Amundsen, P A, 1999. Resource partitioning between Siberian sculpin (Cottus poecilopus Heckel) and Atlantic salmon parr (Salmo salar L.) in a sub-Arctic river, northern Norway. Ecol. Freshwater Fish 8 (4), 201–208.

Gard, M, 2014. Modelling changes in salmon habitat associated with river channel restoration and flow-induced channel alterations. River Res. Appl. 30 (1), 40–44.

Goldsmith, J, Schwartz, J E, 2017. Variable selection in the functional linear concurrent model. Stat. Med. 36 (14), 2237–2250.

Górecki, T, Smaga, Ł, 2019. fdANOVA: an R software package for analysis of variance for univariate and multivariate functional data. Comput. Statis. 34 (2), 571–597.

Groshens, T, Orth, D, 1993. Transferability of habitat suitability criteria for smallmouth bass, Micropterus dolomieu. Rivers 4 (3), 194–212.

Guay, J, Boisclair, D, Leclerc, M, Lapointe, M, 2003. Assessment of the transferability of biological habitat models for Atlantic salmon parr (Salmo salar). Can. J. Fish. Aquat.Sci. 60 (11), 1398–1408.

Guay, J, 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). Can. J. Fish. Aquat.Sci. 57 (10), 2065–2075.

Hayes, Jand, Jowett, I, 1994. Microhabitat models of large drift-feeding brown trout in three New Zealand rivers. North Am. J. Fisheries Manag. 14 (4), 710–725.

Hedger, R, Dodson, J, Bergeron, Nand, Caron, F, 2004. Quantifying the effectiveness of regional habitat quality index models for predicting densities of juvenile Atlantic salmon (Salmo salar L.). Ecol. Freshwater Fish 13 (4), 266–275.

Hedger, R, Dodson, J, Bourque, J, Bergeron, N, Carbonneau, P, 2006. Improving models of juvenile Atlantic salmon habitat use through high resolution remote sensing. Ecol. Modell. 197 (3), 505–511.

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 12 (2–3), 155–169.

Heggenes, J, Bagliniere, J, Cunjak, R, 1999. Spatial niche variability for young Atlantic salmon (Salmo salar) and brown trout (S. trutta) in heterogeneous streams. Ecol. Freshwater Fish 8 (1), 1–21.

Ivlev, V, 1961. Experimental Ecology of the Feeding of Fishes. Yale Univ. Press, New Haven, CT, p. 302.

Jacobs, J, 1974. Quantitative measurement of food selection. Oecologia 14 (4), 413–417.

Jacques, J, Preda, C, 2014. Functional data clustering: a survey. Adv. Data Anal. Classific. 8 (3), 231–255.

Jorde, K, Schneider, M, Peter, A, Zoellner, F, 2001. Fuzzy based models for the evaluation of fish habitat quality and instream flow assessment. In: Proceedings of the 3rd international symposium on environmental hydraulics. pp. 27–28.

Jowett, I G, Hayes, J W and, Duncan, M J, 2008. A guide to instream habitat survey methods and analysis. NIWA Sci. Technol. Series, 54.

Lamouroux, N, Capra, H, Pouilly, M, Souchon, Y, 1999. Fish habitat preferences in large streams of southern France. Freshwater Biol. 42 (4), 673–687.

Larabi, S, St-Hilaire, A, Chebana, F, Latraverse, M, 2017. Multi-Criteria Process-Based Calibration Using Functional Data Analysis to Improve Hydrological Model Realism. Water Resour. Manage. 1–17.

Leclerc, M, Boudreau, P, Bechara, J A, Belzile, L, 1996. Numerical method for modelling spawning habitat dynamics of landlocked salmon, Salmo salar. Regulated Rivers 12 (2–3), 273–285.

Leclerc, M, Boudreault, A, Bechara, T A, Corfa, G, 1995. Two-dimensional hydrodynamic modeling: a neglected tool in the instream flow incremental methodology. Trans. Am. Fish. Soc. 124 (5), 645–662.

Leclerc, M, Saint-Hilaire, A, Bechara, J, 2003. State-of-the-art and perspectives of habitat modelling for determining conservation flows. Canadian Water Resources Journal 28 (2), 135–151.

Leftwich, K N, Angermeier, P L, Dolloff, C A, 1997. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Trans. Am. Fish. Soc. 126 (5), 725–734.

Mäki-Petäys, A, Huusko, A, Erkinaro, J, Muotka, T, 2002. Transferability of habitat suitability criteria of juvenile Atlantic salmon (Salmo salar). Can. J. Fish. Aquat.Sci. 59 (2), 218–228.

Manly, B, McDonald, L, Thomas, D L, McDonald, T L, Erickson, W P, 2007. Resource selection by animals: statistical design and analysis for field studies. Springer Science and Business Media.

Masselot, P, Chebana, F, Ouarda, T B, Bélanger, D, St-Hilaire, A, Gosselin, P, 2018. A new look at weather-related health impacts through functional regression. Sci. Rep. 8 (1), 1–9.

Masselot, P, Dabo-Niang, S, Chebana, F, Ouarda, T B, 2016. Streamflow forecasting using functional regression. J. Hydrol. 538, 754–766.

Mocq, J, St-Hilaire, A, Cunjak, R A, 2013. Assessment of Atlantic salmon (Salmo salar) habitat quality and its uncertainty using a multiple-expert fuzzy model applied to the Romaine River (Canada). Ecol. Modell. 265, 14–25.

Mocq, J, St-Hilaire, A, Cunjak, R A, 2018. Do habitat measurements in the vicinity of Atlantic salmon (Salmo salar) parr matter? Fisheries Manag. Ecol. 25 (1), 22–30.

Moir, H, Gibbins, C N, Soulsby, C, Youngson, A, 2005. PHABSIM modelling of Atlantic salmon spawning habitat in an upland stream: testing the influence of habitat suitability indices on model output. River Res. Appl. 21 (9), 1021–1034.

Morantz, D, Sweeney, R, Shirvell, C, Longard, D, 1987. Selection of microhabitat in summer by juvenile Atlantic salmon (Salmo salar). Can. J. Fish. Aquat.Sci. 44 (1), 120–129.

Morris, J S, 2015. Functional regression. Ann. Rev. Statis. Its Appl. 2, 321–359.

Naman, S M, Rosenfeld, J S, Neuswanger, J R, Enders, E C, Hayes, J W, Goodwin, E O, Jowett, I G, Eaton, B C, 2020. Bioenergetic Habitat Suitability Curves for Instream Flow Modeling: Introducing User-Friendly Software and its Potential Applications. Fisheries 45 (11), 605–613.

Parasiewicz, P, 2001. MesoHABSIM: A concept for application of instream flow models in river restoration planning. Fisheries 26 (9), 6–13.

Poff, N L, Zimmerman, J K, 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biol. 55 (1), 194–205.

Quenouille, M H, 1949. Approximate tests of correlation in time-series. J. R. Statis. Soc.. Series B (Methodological) 11 (1), 68–84.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Railsback, S F, 2016. Why it is time to put PHABSIM out to pasture. Fisheries 41 (12), 720–725.

Ramsay, J, 1982. When the data are functions. Psychometrika 47 (4), 379–396.

Ramsay, J O, Silverman, B W, 2005. Functional Data Analysis. Springer Science and Business Media, New York, NY, United States.

Rosenfeld, J, Beecher, H, Ptolemy, R, 2016. Developing bioenergetic-based habitat suitability curves for instream flow models. North Am. J. Fisheries Manag. 36 (5), 1205–1219.

Rosenfeld, J S, Campbell, K, Leung, E S, Bernhardt, J, Post, J, 2011. Habitat effects on depth and velocity frequency distributions: Implications for modeling hydraulic variation and fish habitat suitability in streams. Geomorphology 130 (3–4), 127–135.

Rudemo, M, 1982. Empirical choice of histograms and kernel density estimators. Scand. J. Stat. 65–78.

Scott, D, Shirvell, C, 1987. A critique of the instream flow incremental methodology and observations on flow determination in New Zealand, Regulated streams. Springer, pp. 27–43.

Smith J, Bérubé F and Bergeron N (2005) A field application of particle image velocimetry (PIV) for the measurement of surface flow velocities in aquatic habitat studies. *26th Canadian Symposium on Remote Sensing.*

Stewart-Koster, B, Olden, J D, Gido, K B, 2014. Quantifying flow–ecology relationships with functional linear models. Hydrol. Sci. J. 59 (3–4), 629–644.

Strakosh, T, Neumann, R, Jacobson, R, 2003. Development and assessment of habitat suitability criteria for adult brown trout in southern New England rivers. Ecol. Freshwater Fish 12 (4), 265–274.

Ternynck, C, Ben Alaya, M A, Chebana, F, Dabo-Niang, S, Ouarda, T B, 2016. Streamflow hydrograph classification using functional data analysis. J. Hydrometeorol. 17 (1), 327–344.

Tharme, R E, 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. River Res. Appl. 19 (5–6), 397–441.

Tukey JW (1977) *Exploratory Data Analysis.*Reading, Mass.

Vadas, R L, Orth, D J, 2001. Formulation of habitat suitability models for stream fish guilds: do the standard methods work? Trans. Am. Fish. Soc. 130 (2), 217–235.

Yen, J D, Thomson, J R, Paganin, D M, Keith, J M and Mac, Nally, R, 2015. Function regression in ecology and evolution: FREE. Methods Ecol. Evol. 6 (1), 17–26.

Yi, Y, Cheng, X, Yang, Z, Wieprecht, S, Zhang, S, Wu, Y, 2017. Evaluating the ecological influence of hydraulic projects: A review of aquatic habitat suitability models. Renewable Sustainable Energy Rev. 68, 748–762.