Effects of spatial scale and choice of statistical model (linear versus tree-based) on determining species–habitat relationships

Sapna Sharma, Pierre Legendre, Daniel Boisclair, and Stéphane Gauthier

Abstract: The choice of spatial scale and modelling technique used to capture species–habitat relationships needs to be considered when ascertaining environmental determinants of habitat quality for species and communities. Fish densities and environmental data were collected at three Laurentian lakes using underwater surveys by four snorkelers collecting fine spatial data acquired through geographic positioning systems. At both fine (20 m) and broad (100 m) spatial scales, tree-based approaches, which incorporated both linear and nonlinear relationships, explained more variation than their linear counterparts. At the finest spatial scale considered (20 m), local environmental conditions, such as habitat structure and heterogeneity, were important determinants of fish habitat selection. At the broadest spatial scale considered (100 m), fish tended to select habitat based on both local environmental features and riparian development. Moran's eigenvector maps further revealed that fish–habitat associations were operating at broader spatial scales than the predefined analytical units, which can be partially attributed to the spatial structure of environmental conditions acting at spatial scales greater than 100 m. This study highlights the importance of evaluating statistical approaches at different spatial scales to identify key determinants of habitat quality for species, ultimately to assess the effects of perturbations on ecosystems.

Résumé : Le choix de l'échelle spatiale et de la technique de modélisation utilisées pour cerner les relations espèce–habitat doit être pris en considération dans l'évaluation des déterminants environnementaux de la qualité de l'habitat pour les espèces et les communautés. Des densités de poissons et des données environnementales ont été mesurées pour trois lacs laurentiens dans le cadre de levés sous-marins par quatre plongeurs en apnée qui ont recueilli des données de haute résolution spatiale à l'aide de systèmes de positionnement global (GPS). Aux échelles spatiales tant fine (20 m) que plus grossière (100 m), les approches arborescentes, qui intègrent des relations linéaires et non linéaires, expliquaient une plus grande partie de la variation que leurs pendants linéaires. À l'échelle spatiale étudiée la plus fine (20 m), les conditions environnementales locales, comme la structure et l'hétérogénéité de l'habitat, étaient des déterminants importants de la sélection de l'habitat par le poisson. À l'échelle la plus grossière (100 m), les poissons avaient tendance à choisir leur habitat en fonction de caractéristiques environnementales locales et de l'aménagement des rives. Les cartes de vecteurs propres de Moran (MEM) ont en outre révélé que les associations poisson–habitat s'exprimaient à des échelles spatiales plus larges que les unités analytiques prédéfinies, ce qui peut être partiellement attribuable à la structure spatiale des conditions environnementales opérant à des échelles spatiales supérieures à 100 m. L'étude met en relief l'importance d'évaluer les approches statistiques à différentes échelles spatiales pour cerner les déterminants clés de la qualité de l'habitat pour les espèces afin, ultimement, d'évaluer les effets de perturbations sur les écosystèmes.

[Traduit par la Rédaction]

Introduction

Habitats can be altered by anthropogenic means, such as resource exploitation and habitat destruction, and can have wide-ranging ecological repercussions, including alterations in habitat availability. Numerous studies have shown that species across many taxa select habitat based on environmental conditions. For example, studies on arthropods (e.g., Schaffers et al. 2008), butterflies and birds (e.g., Storch et al. 2003), woodland caribou (e.g., Leroux et al. 2007), white-tailed deer (e.g., Plante et al. 2004), and freshwater fish communities (e.g., Sharma and Jackson 2007) have demonstrated this pattern. More specifically for freshwater fish communities, environmental variables describing habitat heterogeneity and

S. Sharma*, P. Legendre, and D. Boisclair. Département de sciences biologiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, QC H3C 3J7, Canada.
S. Gauthier†. National Institute of Water and Atmospheric Research, Private Bag 14-901 Kilbirnie, Wellington, New Zealand.

Corresponding author: Sapna Sharma (e-mail: sapna.sharma23@gmail.com).

*Present address: Department of Biology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.
†Present address: Institute of Ocean Sciences, 9860 West Saanich Road, P.O. Box 6000, Sidney, BC V8L 4B2, Canada.

2096

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

riparian development have been used to explain the ecological niche of fish communities. Complex habitat structure and heterogeneity have been shown to increase diversity in fish communities (MacRae and Jackson 2001; Mayo and Jackson 2006), whereas decreased integrity of the riparian zone has been negatively related to diversity in the fish community (Meador and Goldstein 2003).

Species respond to environmental conditions and select habitat at different spatial scales. In this study, spatial scale is defined as the size of the units used during statistical analyses (analytical units) that correspond to the sampling sites at which species and habitat data are collected (Dungan et al. 2002; Fortin and Dale 2005; Brind'Amour and Boisclair 2006). Depending upon spatial scale, apparent species–habitat associations may vary, thereby providing more detailed understanding of the underlying biological processes driving species distributions. Numerous studies highlight the importance of examining species–habitat relationships at different spatial scales to detect patterns of habitat selection (e.g., Graf et al. 2005; Huettmann and Diamond 2006; Guisan et al. 2007a). For example, Plante et al. (2004) related white-tailed deer (*Odocoileus virginianus*) distribution to landscape features at both the 500 m and 1 km scale and found that deer were selecting habitat features at the broader spatial scale. Thompson et al. (2001) examined habitat selection of longnose dace (*Rhinichthys cataractae*) at three hierarchical spatial scales. Significant relationships between longnose dace and their abiotic and biotic environment were only found at the secondary and tertiary (finer) scales (Thompson et al. 2001). Brind'Amour et al. (2005) hypothesized that species' habitat associations can be described at different spatial scales. For example, a generalist fish species that inhabits a broad ecological range should use a wide range of habitat characteristics spanning fine to broad spatial conditions owing to their ability to utilize different types of habitats. Conversely, a specialist fish species that typically uses a narrow ecological range can be associated with very few spatial conditions, as specialist species tend to be adapted to specific habitat characteristics that would only be found at a few spatial scales. The distribution of species can be modelled by eigenfunctions corresponding to the habitat characterizing its ecological niche to identify the spatial scales at which the species was responding to its environmental conditions. Therefore, the distribution of the specialist species will be linked to a small range of spatial scales, represented by the few eigenfunctions modelling it (Brind'Amour et al. 2005). Thus, the choice of spatial scale may yield different types of information as to the habitat features that species are selecting. The advent of spatial statistical methods, such as Moran's eigenvector maps (MEMs; Borcard and Legendre 2002; Borcard et al. 2004; Dray et al. 2006), permits quantification of the size and relative importance of spatial scales acting on environmental conditions and community composition by providing a more objective approach to which spatial scales are important to species communities and their habitat. The added value of including the spatial representation of MEM variables is the ability to quantify the spatial dependency at multiple spatial scales.

Species habitat selection can be characterized in several ways. If species are highly mobile, resource selection functions are useful, as species occurrence is designated by available and locations used by species and are commonly used in wildlife ecology (e.g., Boyce et al. 2002; Manly et al. 2002; Lele and Keim 2006). For species that are not as highly mobile, ecological niche models have often been used to characterize habitat use by species and communities based on a suite of biologically relevant environmental conditions (e.g., Ferrier et al. 2007; Araújo and New 2007; Elith and Leathwick 2009).

The ability to develop ecological niche models with high explanatory power is imperative to understanding the importance of environmental conditions and habitat selection of species in an effort to ultimately assess the effects of perturbations and develop conservation and adaptive management strategies for ecosystems. As statistical approaches differ in their ability to model relationships, an evaluation of different statistical approaches can provide insight into which approach is most appropriate for the biological question being asked at both the species and community levels (Guisan and Zimmermann 2000; Elith et al. 2006; Sharma and Jackson 2008). Numerous studies have compared a suite of statistical approaches (e.g., Elith et al. 2006; Lawler et al. 2006; Prasad et al. 2006; Cutler et al. 2007; Guisan et al. 2007b; Peters et al. 2007; Elith and Leathwick 2009; Knudby et al. 2010; Oppel and Huettmann 2010; Drew et al. 2011; Evans et al. 2011; Hardy et al. 2011), including linear models, generalized linear models, generalized additive models, classification and regression trees (Breiman et al. 1984; De'ath and Fabricius 2000), multivariate regression trees (De'ath 2002), and other models such as Maxent (Phillips et al. 2006; Phillips and Dudik 2008), multiadaptive regression splines (MARS; Friedman 2001; Leathwick et al. 2005), boosted regression trees (De'ath 2007; Buhlmann 2004; Elith et al. 2008), bagging regression trees (Breiman 1996; Buhlmann 2004), and random forest methods (Breiman 2001; Evans et al. 2011). Typically in comparisons where tree-based approaches were included, they generally performed better than many of the other methods. Among tree-based approaches, random forests are considered to be superior (e.g., Lawler et al. 2006; Prasad et al. 2006; Knudby et al. 2010). Random forests have been recently extended to consider a multivariate version (Segal and Xiao 2011) and gradients (Ellis et al. 2012). However, further studies on statistical comparisons are warranted to show which statistical approaches are best to use for species occurrence and community composition data sets.

The primary objective of our study was to ascertain which environmental determinants of habitat quality are important to fish species and communities by considering two methodological aspects: (*i*) identifying the spatial scale at which environmental processes are acting on fish species and communities and (*ii*) partitioning the linear and nonlinear relationships acting on fish species and communities. The analyses were performed at both the species and community levels, as conservation strategies are often developed for both target species and native communities. More specifically, we extended the use of MEMs to identify the spatial scales at which environmental variables are structuring fish communities without using predefined scales and used a variation partitioning framework to attribute variation to linear and nonlinear components when modelling community–habitat relationships using multivariate approaches. This study focused on modelling the relationships between littoral fish species and their environment at fine and broad spatial scales using a data set typical of those collected by aquatic ecologists.

Table 1. Occurrence of fish (number of sections in which species is present) and maximum and mean densities of fish observed in the littoral zones of Purvis Lake, Rond Lake, and Violon Lake in July 2005 in sampling units consisting of sections 20 m long.

Fish species	Occurrence	Maximum	Mean
Lake Purvis (123 sections)			
Pumpkinseed (*Lepomis gibbosus*)	114	1.11	0.26
Rock bass (*Ambloplites rupestris*)	66	0.06	0.01
Smallmouth bass (*Micropterus dolomieu*)	98	0.08	0.02
Yellow perch (*Perca flavescens*)	34	0.04	0.003
Lake Rond (81 sections)			
Banded killifish (*Fundulus diaphanus*)	75	3.66	0.41
Bluntnose minnow (*Pimephales notatus*)	21	1.00	0.15
Goldfish (*Carassius auratus*)	18	0.02	0.002
Pumpkinseed (*Lepomis gibbosus*)	79	0.54	0.15
Smallmouth bass (*Micropterus dolomieu*)	67	0.07	0.01
Lake Violon (116 sections)			
Creek chub (*Semotilus atromaculatus*)	4	0.05	0.001
Fathead minnow (*Pimephales promelas*)	16	0.01	0.001
Golden shiner (*Notemigonus crysoleucas*)	54	0.26	0.02
Pearl dace (*Margariscus margarita*)	66	0.26	0.01
Pumpkinseed (*Lepomis gibbosus*)	104	1.05	0.09
Walleye (*Sander vitreus*)	4	0.02	0.0002
White sucker (*Catostomus commersonii*)	20	0.03	0.001

Note: The minimum densities of all fish in all lakes was zero.

In this study, we limited our comparison to linear (multiple linear regression and redundancy analysis) and tree-based (regression tree and multivariate regression trees) models at different spatial scales to identify important environmental variables structuring fish species and communities. Linear approaches are often used to relate environmental conditions to species occurrence (Sharma et al. 2008). Tree-based approaches can be beneficial to use with data sets that exhibit linear and nonlinear relationships between predictor variables, high-order interactions, and multicollinearity, and they are generally faster and more accurate than traditional linear approaches (Breiman et al. 1984; De'ath and Fabricius 2000). We limited our comparison to the aforementioned approaches for several reasons: (*i*) multivariate versions were available for both linear and tree-based methods; (*ii*) direct identification of important predictor variables was possible; (*iii*) to permit quantification of the importance of fine to broad spatial scales using a variation partitioning framework in conjunction with MEMs; and (*iv*) to allow us to quantify and partition the variation attributed to linear and nonlinear processes acting on fish communities, thereby presenting a unique contribution to the literature.

Methods

Study lakes

Our study focused on the littoral zone of three lakes of comparable sizes located in the Quebec Laurentian region: Purvis Lake (45.99°N, 74.09°W), Rond Lake (45.95°N, 74.14°W), and Violon Lake (45.94°N, 74.09°W). Fish species diversity does not tend to be high in small Laurentian lakes.

There were four and five fish species found in the littoral regions of Purvis and Rond Lakes, respectively, in July 2005. The fish community in Violon Lake comprised seven species (Table 1).

Purvis Lake has a surface area of 19.1 ha, a perimeter of 2.4 km, and a mean depth of 7.6 m. It is a dystrophic lake with a single tributary. Human development around the perimeter of Purvis Lake is intermediate, with 27 cottages. The littoral zone of Purvis Lake is characterized by a diversity of habitats, including sandy beaches, rocky substrate, woody debris, and patches of macrophytes. Rond Lake has a surface area of 16.6 ha, a perimeter of 1.6 km, and a mean depth of 7.2 m. Rond Lake is highly developed and presently hosts 314 cottages. The littoral zone of Rond Lake is relatively homogenous and dominated by weed beds. The surface area of Violon Lake is 15.5 ha, the perimeter is 2.5 km, and the mean depth is 7.3 m. Of the three lakes, Violon Lake is the least developed and is home to 3 cottages. Violon Lake is characterized by the presence of numerous dead trees and high coverage of woody debris and macrophytes, which are commonly removed by cottage owners (Table 2).

Characterization of the fish community

The littoral zone fish community was described by conducting three surveys in each lake. The three surveys were completed during the day between 1000 and 1600 h of July 2005. Please see Brind'Amour and Boisclair (2006) for a complete description of the sampling methodology. The surveys conducted in this study were a modified form of the visual approach used by Brind'Amour and Boisclair (2006). Our

2098

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Table 2. Mean environmental conditions observed in the littoral zones of Purvis Lake, Rond Lake, and Violon Lake, in July 2005 in sampling units consisting of sections 20 m long.

Variable	Purvis Lake	Rond Lake	Violon Lake
Depth at 10 m from shoreline (m)	2.9	—	4.8
Bottom slope (°)	16	10.7	24.9
Clay substrate (%)	0	0.7	0.04
Silt substrate (%)	59.5	73.1	58.7
Sand substrate (%)	5.2	19.4	2.7
Gravel substrate (%)	0.3	2.4	0.5
Pebble substrate (%)	0.3	0.1	0.1
Cobble substrate (%)	0.6	0.1	3.7
Boulder substrate (%)	10.6	1.6	16.2
Metric boulder substrate (%)	12.9	2.6	14.9
Bedrock substrate (%)	10.6	0	3.1
Macrophyte coverage (%)	38.6	67.9	10.6
Woody debris (%)	23.2	0.4	27.3
Abundance of dead tree trunks	3.26	0.1	26.3
Shoreline vegetation[a]	3.3	—	3.9

[a]Shoreline vegetation key: 1, lawn grass; 2, natural vegetation; 3, wooded area; 4, forest.

approach involved four snorkellers (instead of the two snorkellers in Brind'Amour and Boisclair 2006) trained and calibrated to identify, count, and estimate the length of the fish observed during visual surveys. From a common starting point, two snorkellers surveyed the perimeter of the lake clockwise and the other two snorkellers surveyed the perimeter of the lake counterclockwise, eventually meeting after surveying approximately 50% of the perimeter of the lake. For each pair of snorkellers, one swam approximately 1 m from shore and one swam near the 2.5 m depth isobath (limit to discriminate and identify fish during the day and night in the study lakes). In contrast to the approach used by Brind'Amour and Boisclair (2006), who used painted rocks to set the limits of predefined sampling sites, in the present study all snorkellers carried a geographic positioning system (GPS) (Garmin model; ±5 m under prevailing field conditions) that recorded their position at 1 s intervals. Snorkellers recorded the species, the number, and the length of the fish (classes of 5 cm total length) they observed together with the time of observation (hh:mm) on polyvinyl chloride tubes they wore on their forearms. Snorkellers also collected information required to estimate the volume they sampled by estimating variations of the width and the depth of the area they surveyed (at an accuracy of 25 cm based on distance to shoreline and visibility). Combining the data of the position of the snorkellers and the time noted by the GPS together with the data of the fish observed and the time of observation noted by the snorkellers permitted the estimation of the spatial distribution of fish around the complete perimeters of the lakes.

We divided the perimeter of each lake into a series of contiguous 20 m long sections. The length of 20 m sections was selected as the finest spatial scale on which models could be developed to summarize species–habitat relationships. Given the precision of the GPS used (±5 m), the maximum difference between the time at which the position of snor-

kellers and that of fish were recorded (1 min), and the mean swimming speed of snorkellers (5 m·min⁻¹; e.g., half a perimeter of 2.5 km lake in 4 h), the potential error of the spatial position of fish at any given point of the perimeter of the lakes is taken as ±10 m. This means that at the limit of a given 20 m section, the probability of assigning fish to the proper 20 m section is 50%. This probability reaches 75% at 5 m from the limit of a 20 m section and 100% in the centre of this section. Taken together, the probability of assigning fish to the 20 m section in which they were effectively observed is 70%. This value is taken as the worst-case scenario associated with fish position because (*i*) it assumes that the error imputable to the GPS and that associated with the time difference between the positioning of the snorkeler and the fish are always perfectly additive and (*ii*) it assumes that the swimming speed of the snorkeler is constant, while it is certainly slower (and hence, less subjected to an error due to the difference in time of the positioning of the snorkeler and the fish) when snorkellers encounter fish than when they go through a fishless area.

Characterization of environmental conditions

Corresponding to the characterization of the fish community, a suite of local and riparian environmental data were recorded for each 20 m section over the complete perimeter of the study lakes. Local habitat variables were noted and georeferenced by snorkellers within a few days of sampling for fish. Habitat variables designated as local were the slope at the bottom of the lake (estimated using the depth at a point 20 m from shore), the composition of the substrate (percent cover by nine size categories of substrate as identified by Bouchard and Boisclair (2008) from clay to bedrock), the percent cover of macrophytes, and the number of dead trees with trunk diameters greater than 15 cm. Riparian habitat variables were estimated from a boat following the shoreline. Riparian habitat variables were all categorical variables: for example, the slope of the shoreline (1 = 0°–30°, 2 = 30°–60°, 3 = 60°–90°), and the use of riparian zone such as the presence or absence of cottages, walls, platforms, docks, or vegetation (i.e., lawn grass, natural vegetation, wooded area, and forest).

Data analyses

Data were analysed at two spatial scales (analytical units on which statistical analyses was performed): 20 and 100 m. The largest analytical unit was an amalgamation of data in five elementary units, thus a 100 m section. Amalgamation of data was completed as follows. Fish counts and sampling volumes were summed, and the ratio of the total number of fish in a sampling volume was the density of fish in a 100 m section. Values for the slope at the bottom of the lake, composition of substrate, and the percent cover of macrophytes were averaged across the five elementary units to obtain mean conditions in a 100 m section. The number of dead trees with trunk diameters greater than 15 cm was summed for a 100 m section. Finally, the mode value of the slope of the shoreline and the use of the riparian zone were taken for each 100 m section.

For each sampling site we calculated the density of each fish species per volume sampled. Volume was calculated for each sampling site using the length of the transect, the width of the sampling area, and mean lake depth at that particular sampling site. For univariate analysis, the square-root transformation was used to attempt to achieve normality of the residuals of the

response variable prior to analysis. For multivariate analyses, fish density data were subjected to the Hellinger transformation. The transformation consists of expressing each fish density as a proportion of the sum of all densities in the analytical unit and taking the square root of the resulting value (Legendre and Gallagher 2001). The square-root portion of the transformation decreases the importance of the most abundant species. This transformation is recommended for use in linear ordinations (Rao 1995; Legendre and Legendre 1998; Legendre and Gallagher 2001). All data analyses were performed in the R language environment (R Development Core Team 2010).

Performance of environmental variables was assessed based on the proportion of the response variable's variation explained in the data set by each statistical approach (R^2). Since R^2 as an estimator of the proportion of variation explained is biased and increases with the number of explanatory variables even if they are random, we calculated the adjusted R^2 (R^2_{adj}), which provides unbiased estimates of the variation of the response data explained by the explanatory variables (Ohtani 2000, Peres-Neto et al. 2006). Adjusted R^2 allows a comparison of linear models composed of different numbers of predictor variables and sample sizes in a statistically unbiased manner (Peres-Neto et al. 2006). However, if different transformations of the data or different measures of variation are used, the valid comparisons are between linear or tree-based methods within univariate and multivariate responses at different spatial scales (Anderson-Sprecher 1994).

Linear approaches

Multiple regression assumes that the mean of the response variable can be expressed as a linear combination of known functions of the predictor variables. The "linearity" in linear models is with respect to the unknown parameters, not the predictor variables. The multivariate counterpart we adopt here is redundancy analysis (RDA), which is a canonical analysis approach (Legendre and Legendre 1998). In both cases, the R^2 is the proportion of the species or community variation explained by a linear combination of the environmental variables (Legendre and Legendre 1998). We calculated the adjusted R^2 for the linear approaches as $R^2_{adj} = 1 - (1 - R^2) \times [(n - 1)/(n - m - 1)]$, where m is the number of regressors and n is sample size (Legendre and Legendre 1998, p. 525). For a fair comparison with tree-based models, second- and third-degree polynomials of each quantitative environmental variable were used to model curvilinear species–habitat relationships. The same suite of predictor variables together with their higher degree polynomial terms were all initially used for the univariate and multivariate models. Environmental variables were then selected using a modified forward selection procedure that corrects for highly inflated type I error and overestimated amounts of explained variation, which are classical problems of forward selection (Blanchet et al. 2008). Environmental variables that were significant at an α level of 0.05 based on 999 random permutations were subsequently used in multiple regression and redundancy analysis.

Tree-based approaches

Regression trees and multivariate regression trees were used as representative approaches from the family of tree-based approaches for a statistical comparison with the linear methods. We limited our comparison to regression trees and mul-

tivariate regression trees from the large number of tree-based approaches for several reasons: (*i*) multivariate versions with known ecological applications were available for both linear and tree-based methods; (*ii*) direct identification of important predictor variables was possible; and (*iii*) the ability to quantify and partition the variation attributed to linear and nonlinear processes acting on fish communities was possible. Regression trees and multivariate regression trees iteratively divide data into two homogenous groups along the values of one of the explanatory variables in such a way that they have mutually exclusive memberships and minimize the variation (sum of squares) of the response variable(s) within the two groups. Regression trees can be constructed using continuous and (or) categorical predictor variables (Breiman et al. 1984; De'ath and Fabricius 2000; De'ath 2002). We constructed regression trees using all environmental variables with the constraint that there was a minimum sample size of six in the groups produced at each split. The most parsimonious regression trees were selected by pruning the trees to the level where the complexity parameter minimized the cross-validation error. All regression trees were developed using the rpart package in R, and multivariate regression trees were developed in the mvpart package in R (Ripley 2007; R Development Core Team 2010). The percent variation (R^2) explained by the regression tree using the predictor variables was calculated using $R^2 = 1 - $ relative error. The relative error is the sum, over all groups of a partition level, of the within-group sum of squares, divided by the total sum of squares of the response data.

Variation partitioning of multivariate linear and tree-based approaches

We compared the results of redundancy analyses and multivariate regression trees at both spatial scales by partitioning the variation explained by each method using the same set of predictor variables. This approach summarized the amount of variation in the fish community data that is responding linearly and nonlinearly to environmental conditions.

Spatial–environmental relationships

The distance-based MEM was used to quantify symmetric spatial structure, that is, spatial patterns that do not incorporate an assumption of directionality. MEMs are obtained by eigen decompostion of the product of a matrix of connectivity among the sites, representing the edges or links between sites, by a matrix of edge weights. These MEM variables, or eigenfunctions, are obtained from a spectral decomposition of a truncated distance matrix of the spatial relationships among sampling locations. Distance-based MEM eigenfunctions are constructed by computing the Euclidean distance among the sites (the centre of each 20 m section). In the matrix of Euclidean distances, one finds the smallest distance that maintains connections between all sites along a minimum spanning tree; this value is called the truncation distance. The distances between sites that are larger than the truncation distance are replaced by four times the truncation distance. The modified, truncated distance matrix is referred to as the neighbour matrix. A principal coordinates analysis (PCoA) is conducted on the neighbour matrix. Two-thirds of the eigenvalues produced by the PCoA will be positive. The MEM eigenfunctions modelling positive autocorrelation (i.e., eigenfunctions with Mo-

2100

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Fig. 1. Schematic description of the procedure used to determine the smallest spatial scale modelled by the Moran's eigenvector map variables. One counts the number of like-coloured squares (i.e., black or white) in a patch multiplied by the size of the analytical unit (20 m section). In this hypothetical example of a linearized set of sites, the finest spatial processes are acting at a scale of 60 m.
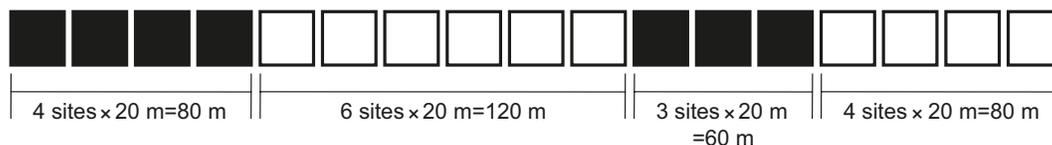
ran's I larger than the expected value) will be referred to as MEM variables and represent the spatial structure in the data set. The MEM variables will be in the form of a series of sine waves with decreasing periods that are orthogonal to one another. Finally, a linear analysis (such as redundancy analysis if the response data are multivariate and multiple linear regression if response data are univariate) is performed on the response data with a set of predictor variables (Appendix A; Borcard and Legendre 2002; Borcard et al. 2004). MEM eigenfunctions describe symmetric spatial structures at all spatial scales that can be expressed by the sampling design. The first MEM variables model broad spatial structures, and subsequent MEM variables represent smaller spatial patterns. The last eigenfunctions accommodate fine-scale spatial structures (see Borcard and Legendre 2002; Borcard et al. 2004; Dray et al. 2006 for details).

To quantify the spatial scale at which the environmental conditions were acting on the fish community, we related four groups of environmental features to the MEM spatial descriptors: lake morphology (based on depth, slope, distance to tributary), riparian development, substrate type, and habitat complexity characteristics (percent cover of macrophytes, woody debris, and branches). We identified the significant MEM variables corresponding to each category of environmental variables using forward selection based on a significance level of 0.05 and 999 random permutations. The MEM variables explaining the most variation for each set of habitat features were plotted spatially. Spatial scale was quantified by counting the minimum number of like-coloured squares in a patch (for example, counting the continuous number of black squares) and multiplying that value by the size of the analytical unit (20 m section; Fig. 1). The spatial representation of MEM variables can quantify the spatial dependency at multiple spatial scales rather than a predetermined analytical unit size.

Results

Linear approaches

At the fine spatial scale (20 m), the variation explained by the fish density models developed using multiple regression generally yielded low explanatory power ranging from 0% to approximately 37%. At the broad spatial scale (100 m), the explained variation of the multiple regression models ranged from 0% to approximately 84% (Table 3). At the fine spatial scale, the multiple regression model explaining pumpkinseed density in Violon Lake suggests that higher densities of pumpkinseed were found in regions with intermediate coverage of boulder and clay substrate and tree trunks (Appendices B–C). At the broad spatial scale, the multiple regression model predicting rock bass density in Violon Lake suggests that

Table 3. Mean explained variation (R^2; bolded values) and the range of R^2 values for all statistical approaches (LR (multiple linear regression); RT (regression tree); RDA (redundancy analysis); and MRT (multivariate regression trees)) used to model fish density–habitat relationships in Purvis, Rond, and Violon lakes at fine (S1: 20 m) and broad (S5: 100 m) spatial scales.

Spatial scale	Adjusted R^2	LR	RT	RDA	MRT
Fine (S1)	Mean	**14.3**	**35.3**	**7.3**	**17.4**
Fine (S1)	Minimum	0	0	3.4	9.9
Fine (S1)	Maximum	36.8	71	12.1	25.8
Broad (S5)	Mean	**49.7**	**68.2**	**14.6**	**42.4**
Broad (S5)	Minimum	0	31	4.7	39.2
Broad (S5)	Maximum	84	95.3	27.5	48.5

higher densities of rock bass were found in regions with high abundances of tree trunks, high percent cover of pebble substrate, and low percent cover of cobble substrate (Appendices B–C).

With respect to the fish community, the RDAs were significant at $p < 0.05$ based on 999 random permutations and explained between 3.4% and 12.1% of the variation in density of the fish community at the fine spatial scale (Table 3). Combinations of substrate type and habitat complexity variables were significant in describing the fish community in the Laurentian lakes (Appendix D). At the broad spatial scale, the RDA was significant at $p < 0.05$ based on 999 random permutations and explained between 4.7% and 27.5% of the variation in fish community density (Table 3). Combinations of substrate type, habitat complexity, and riparian habitat were significant in describing the fish community at the broad spatial scale (Appendix D).

Tree-based approaches

At the fine spatial scale, the explained variation of regression tree models ranged from 0% to 71% (Table 3). For example in Lake Violon, highest densities of pumpkinseed were found in regions with low woody debris and intermediate coverage of boulder substrate, macrophytes, and tree trunks. In general, species densities tended to be highest in sites with high natural habitat heterogeneity (i.e., associated with a variety of habitat types; Appendices B–C). At the broad spatial scale, variation explained by regression tree models ranged from 31% to 95% (Table 3). In Lake Purvis, high densities of rock bass were found in the littoral area characterized by a developed riparian habitat, complex habitat structure, and in substrates with a low percent cover of metric boulder (Fig. 2; Appendices B–C).

Fig. 2. Regression tree showing the important environmental variables explaining rock bass densities at the broad spatial scale in Purvis Lake. Rock bass densities are highest when the abundance of tree trunks is greater than 74.5 and where there are more than 1.5 cottages on a site. Rock bass densities are lowest in the presence of fewer tree trunks, intermediate substrate coverage of metric boulders, and the presence of more than three walls on a site.
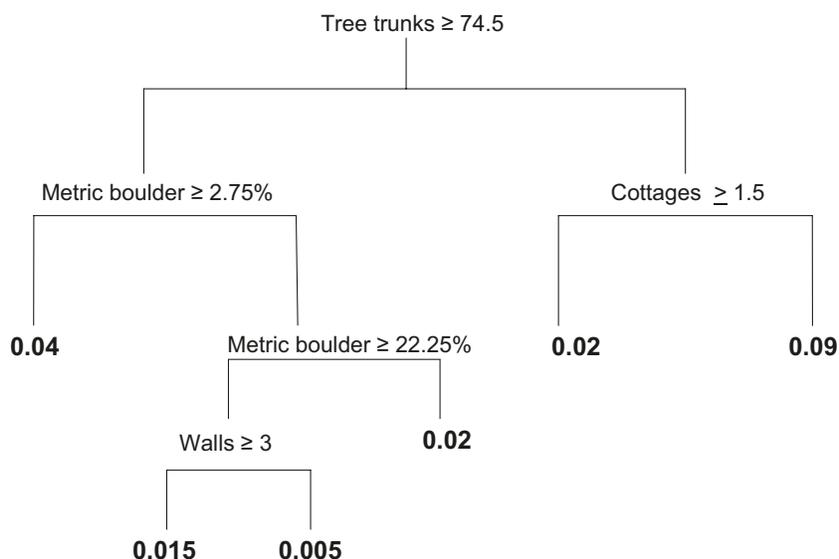
At the fine spatial scale, the multivariate regression tree explained approximately 10%–26% of the variation in the fish community (Table 3). Combinations of substrate and habitat complexity variables were significant in describing the fish community in the Laurentian lakes (Appendix D). At the broad spatial scale, the multivariate regression tree explained approximately 39%–48.5% of the variation in the fish community. Combinations of substrate, habitat complexity, and riparian habitat were significant in describing the fish community at the broad spatial scale (Appendix D).

Comparison of statistical approaches

At the fine and broad spatial scales, the tree-based approaches yielded higher predictive power and outperformed their linear counterparts (Table 3). This indicates that there are interactions between environmental variables and that fish densities are responding in a nonlinear fashion, beyond a third-degree polynomial, to the environmental variables used in this study. Furthermore, the percent variation explained by the community level models was higher at the broad spatial scale, suggesting that the fish in the Laurentian lakes are selecting habitat features and forming community assemblages at the broad spatial scale.

On average, across lakes at the fine and broad spatial scales, 2.2% of the variation is explained uniquely by redundancy analyses. Approximately 43.4% of the variation is shared between redundancy analysis and multivariate regression trees. The remaining 54.3% of the variation is explained solely by multivariate regression trees (Table 4). The shared variation represents linear relationships modelled by both the redundancy analyses and multivariate regression trees. The independent fraction explained solely by multivariate regression trees can be attributed to relationships modelled by multivariate regression trees that are not modelled by redundancy analysis, such as interactions between predictor variables and nonlinear relationships.

Table 4. Percent variation ($R^2_{adjusted}$) explained solely by redundancy analysis, multivariate regression trees, and shared variation between redundancy analysis (RDA) and multivariate regression trees (MRT) for fine and broad spatial scales in Purvis, Rond, and Violon lakes.

	% variation				
Lake	Spatial scale	Unique RDA	Unique MRT	Shared	Total
Purvis	S1	1.9	8.6	1.3	11.8
Purvis	S5	4.9	15.4	15.1	35.4
Rond	S1	−0.8	7.4	18.5	25.1
Rond	S5	−8	18.2	21.3	31.5
Violon	S1	1.2	5	11.4	17.6
Violon	S5	−3.4	22.2	26.4	45.2

Note: All $R^2_{adjusted} \leq 0$ indicate no explained variation. The values resulting from variation partitioning are not R^2, but adjusted R^2, which are unbiased estimates of the amount of variance of a response matrix **Y** explained by explanatory variables X. Adjusted R^2 can take negative values when X explains less of the variation in Y than the same number of random normally distributed predictors would. Hence, negative adjusted R^2 are interpreted as zeros.

The environmental variables identified by the linear and the tree-based approaches tended to be a subset of one another and in other cases were not identical environmental variables (Appendices C and D). For example, at the broad spatial scale in Violon Lake, RDA identified the presence of docks, macrophyte coverage, and silt substrate as significant predictor variables of fish community composition, whereas MRT identified silt and bedrock coverage and the abundance of tree trunks as important environmental conditions structuring fish community composition.

2102

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Table 5. Adjusted percent variation (R^2_{adj}) explained by Moran's eigenvector maps (MEMs) on lake morphology, riparian development, substrate composition, and habitat complexity in Purvis, Rond, and Violon lakes.

Variable	MEM R^2_{adj}			First MEM (R^2_{adj})			First MEM (m)		
	Purvis	Rond	Violon	Purvis	Rond	Violon	Purvis	Rond	Violon
Lake morphology	88.6	90.3	92.3	3 (16.8)	3 (26)	7 (18.8)	1140	140	320
Riparian development	56.1	25.8	5	3 (16.1)	3 (14.7)	15 (2.6)	1140	140	140
Substrate composition	47.7	54.8	44.8	6 (10.7)	2 (8.5)	9 (4.1)	320	400	260
Habitat complexity	84.3	38.1	59	3 (33.1)	1 (4.8)	5 (11.3)	1140	100	420

Note: The first set of columns represents the total variation explained by MEMs on the set of environmental variables (R^2_{adj}). The second set of columns represents the Moran's eigenfunction that explains the greatest amount of variation in the environmental variable, with the percent variation explained expressed in parentheses. The third set of columns represents the spatial scale (m) at which the most important Moran's eigenfunction is acting.

Fig. 3. Map of Lake Purvis (*a*), Lake Violon (*b*), and Lake Rond (*c*) depicting the spatial processes modelled by (*a*) Moran's eigenvector map (MEM) variable 3, (*b*) MEM variable 5, and (*c*) MEM variable 1. Each 20 m section is represented by a square. The black and white squares represent positive and negative values, respectively. The size of the squares is proportional to the forecasted values. To determine the finest spatial scale modelled by the MEM variables, one identifies the number of like-coloured squares in a patch multiplied by the size of the analytical unit (20 m section), as depicted in Fig. 1.
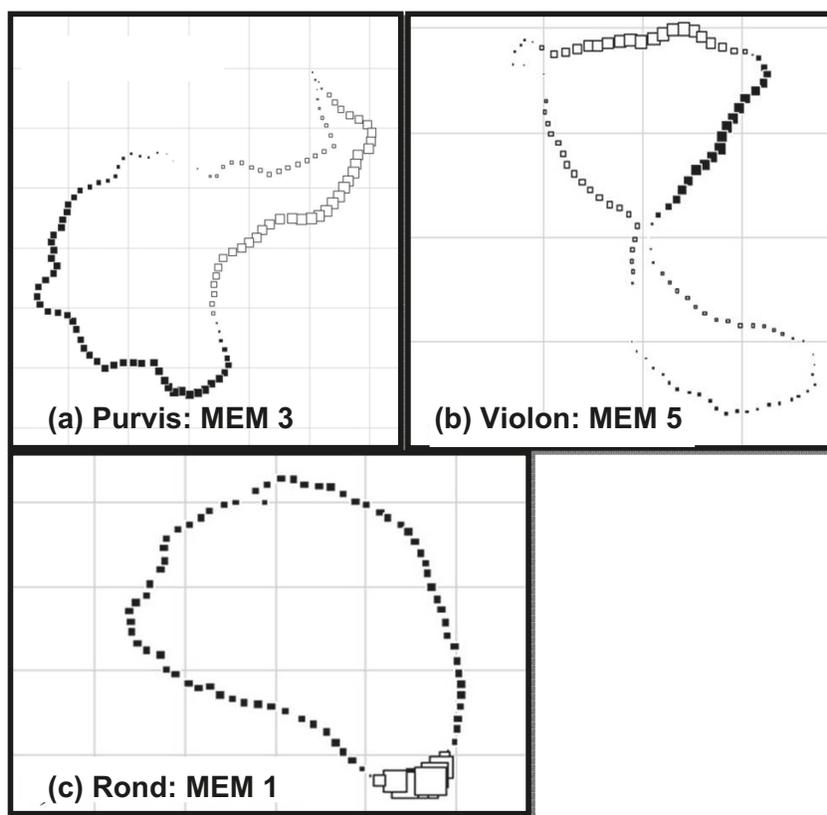
Spatial–environmental relationships

On average, across all lakes, the significant spatial descriptors generated by the MEM analyses explained 90% and 29% of the variation in lake morphology and riparian development, respectively. The MEM variables explaining the most variation in lake morphology and riparian development represented spatial processes acting at between 0.14 and 1.14 km, depending upon the lake. This suggests that variables describing lake morphology and riparian development are acting at a very broad spatial scale in Laurentian lakes (Table 5; Fig. 3).

On average, the significant spatial descriptors generated by the MEM analyses explained 49% of the variation in substrate composition. The MEM variables explaining the most variation in substrate composition represented an intermediate spatial scale acting at the range of 260–400 m (Table 5; Fig. 2). This suggests that substrate conditions tend to be spatially structured at an intermediate spatial scale if habitat is heterogenous, although this scale is larger than the broad (100 m) analytical unit used to model species–habitat relationships.

Finally, the significant spatial descriptors generated by the MEM analyses explained 60.5% of the variation in habitat

complexity features. In Purvis and Violon lakes, the MEM variables describing spatial processes acting at 1.14 and 0.42 km, respectively, explained the most variation in habitat complexity. In Rond Lake, MEM variable 1 explained the most variation in habitat complexity, at least at the 100 m scale, although there were only two spatially structured patches of habitat in the lake (Table 5; Fig. 2). This suggests that habitat complexity features act at a range of spatial scales depending upon the features of each lake.

Discussion

Comparison of statistical approaches

The differential structural properties of a data set require the use of statistical approaches that best capture the response in the data set whilst providing information on the important environmental determinants structuring species' distributions and communities. The comparison of linear versus tree-based statistical approaches at the species and community levels provided some valuable insights into the relationships between species density and habitat. Linear methods are traditionally the most popular approaches used (Sharma et al. 2008). However, we found that at both spatial scales, the tree-based approaches had higher predictive power than their linear counterparts as they explained more variation in species–habitat relationships. Variation partitioning analyses showed that multivariate regression trees captured almost all of the linear variation explained by redundancy analyses, in addition to nonlinear relationships between environmental conditions and fish density. Generally, we found that the environmental variables identified by the linear and tree-based approaches tended to be similar, but not identical. We hypothesize that the inclusion of a subset of environmental conditions in linear versus tree-based methods could be as a result of (*i*) the inclusion of only the environmental conditions exhibiting the strongest relationship with fish densities, (*ii*) the higher explanatory power of regression tree models, and (*iii*) the ability of tree-based approaches to capture both linear and nonlinear relationships.

Tree-based approaches can have high predictive abilities for data sets that exhibit linear and nonlinear relationships between predictor variables, high-order interactions, and multi-collinearity (Breiman et al. 1984; De'ath and Fabricius 2000; De'ath 2002). Tree-based models produce discontinuous changes at certain points along the predictor variables and identify high-order local interactions that the linear-based approaches used in this study do not appear to accommodate. Regression trees provide clear graphical interpretations of the predictor variables and the thresholds required to attain mean densities of species and community assemblages, thereby providing insight into the interpretation of ecological patterns (De'ath and Fabricius 2000). However, in some cases the output from regression trees can be unstable. The regression trees developed in our study certainly provided more ecological information as to which environmental variables are explaining the variation in fish densities. A potential drawback of tree-based approaches can be the possibility of over-fitting a model (Sharma et al. 2008). However, selecting the most parsimonious regression trees by pruning the trees to the level where the complexity parameter minimized the cross-validation error reduced that possibility in our study.

Effect of spatial scale on species–habitat relationships

For all statistical approaches, there was greater explained variation in species–habitat relationships at the broad spatial scale. Species appear to be selecting both local and riparian habitat features at broader spatial scales and only local habitat features at the finer spatial scale. This can be attributed to the broad spatial processes structuring the environmental variables used to generate the models as revealed by MEMs. Thus, a broader spatial scale may more accurately reflect the ecological processes acting on the fish community (Cooper et al. 1998). Furthermore, it is difficult to ascertain community assemblage – environmental relationships at an analytical unit of a 20 m section, as species may not be forming community assemblages at such fine spatial scales. At the broad spatial scale, however, it is possible to identify and predict community assemblage – environmental relationships with a higher degree of power. As such, we conclude that fish in these Laurentian lakes are selecting habitat features and forming community assemblages at the broad spatial scale.

Spatial processes can be easier to detect at a broader spatial scale (larger analytical unit size) in part owing to statistical artefacts. Data in larger analytical units exhibit lower variance due to a reduction in the spread and skewness of data points (Bellehumeur et al. 1997; Rossi and Nuutinen 2004) and removal of fine-scale variation in the study (i.e., the beginning of the variogram (Bellehumeur and Legendre 1997; Bellehumeur et al. 1997; Plante et al. 2004)) illustrated by Rossi and Nuutinen (2004). Further investigation of spatial–environmental relationships using MEM analyses in the study lakes revealed, however, that the environmental features used to model fish density were structured at spatial scales larger than a 100 m section. Identifying the size of the spatial scale at which environmental conditions are operating would not have been quantifiable prior to the advent of spatial statistical methods (Borcard and Legendre 2002; Dray et al. 2006). For example, habitat features describing lake morphology and riparian development tended to be predominately spatially structured at very broad spatial scales (e.g., >1.1 km in Purvis Lake). Conversely, substrate composition features tended to be spatially structured at intermediate spatial scales, whereas habitat complexity features were spatially structured at a range of spatial scales depending upon habitat heterogeneity in the lake, yet at a scale still greater than the largest analytical unit used in the study. The analyses suggested that the spatial structure of the environmental characteristics may be a strong contributor to fish selecting habitat at broader spatial scales, rather than purely due to a statistical artefact (Brind'Amour et al. 2005).

It is most likely that spatial correlation (sensu stricto) is present in the environmental variables, but not in the fish data if the observers were careful not to count the same individual fish in two adjacent sections. There may still be spatial structure ("spatial correlation", not "spatial autocorrelation") in the fish data, but that will be due to environmental control over the species distributions, an effect known as induced spatial dependence. However, simulation studies have shown that when spatial correlation is present in only one of the two variables under study (response, explanatory), tests of significance have correct levels of type I error (Legendre et al. 2002). Therefore, if spatial correlation was present in both the response and explanatory variable(s) using MEM as covariables, spatial

correlation would be effectively corrected for in the results of the significance test (Peres-Neto and Legendre 2010).

Littoral fish assemblages

The fish–habitat relationships ascertained in this study reveal the environmental conditions and the interactions between them that fish are selecting in Laurentian lakes. For example, pumpkinseed was found in all the study lakes at high densities. In Rond Lake, at the 20 m spatial scale, more variation in pumpkinseed density was explained by the regression tree at the 100 m spatial scale (90%) than at the 20 m spatial scale (51.6%). Generally, pumpkinseed was associated with silt substrate, which is typically positively associated with macrophyte and vegetation growth. Pumpkinseed are known to inhabit regions with submerged vegetation and a gradient of substrate types, often to accommodate their planktivorous diet (Scott and Crossman 1973; Robinson et al. 1993). The highest amount of variation explained by regression tree models for rock bass densities was in Purvis Lake at the 100 m spatial scale (95.3%) compared with 60.6% variation explained at the 20 m spatial scale. The highest densities of rock bass were found in the littoral area characterized by a developed shoreline and in substrates consisting of boulder and silt (Keast et al. 1978).

We found that yellow perch preferred shallow regions of the littoral zone with a higher percent cover of fine substrate, which is typically associated with aquatic vegetation. Our findings correspond to the habitat preferences described by Kitchell et al. (1977) and Keast et al. (1978), who found that yellow perch preferred sandy regions in the presence of aquatic vegetation in the littoral zone (Kitchell et al. 1977; Keast et al. 1978). The highest densities of smallmouth bass were found in shallow regions of the littoral zone with moderate percent cover of metric boulder substrate, as they prefer heterogenous habitats provided by rocks (Scott and Crossman 1973). Highest densities of goldfish were positively related to the number of cottages, coverage of macrophytes, and abundance of tree trunks. Goldfish are often found in water bodies with high growth of aquatic plants (Scott and Crossman 1973), are associated with human presence, and continue to be released into watersheds by human-mediated means through direct stocking, fish hatcheries, aquariums, or ornamental ponds (Mills et al. 1993). In Violon Lake, walleye densities were positively related to abundance of tree trunks and negatively related to high coverage of metric boulder. Adult walleye prefer extensive littoral areas of gravel or rubble on which to spawn (McMahon et al. 1984), thereby avoiding areas with larger sized substrate. Furthermore, adult walleye are negatively phototaxic and during the day prefer logs or submerged vegetation (McMahon et al. 1984). Thus, the overall habitat preferences exhibited by the fish in the littoral zones of the study Laurentian lakes correspond to known habitat preferences from the literature.

In Purvis, Rond, and Violon lakes, fish densities were related to both local and riparian environmental features, similar to the findings of Brind'Amour and Boisclair (2006), who also found that both intrinsic (i.e., within lake) and extrinsic (i.e., outside of lake) environmental variables were significant contributors to fish habitat models. However, we also found that at the fine spatial scale, fish densities were primarily determined by local habitat variables. Environmental variables describing habitat structure and heterogeneity, such as woody debris, substrate, and vegetation, have been found to be strong contributors influencing fish community composition (Mayo and Jackson 2006). At the broad spatial scale, local environmental features such as percent cover of woody debris, silt and boulder substrates, and riparian environmental features such as riparian development were significant determinants of fish community densities. This supports the inclusion of both local and riparian environmental data in species–habitat models to improve our understanding of the nature of species–environment relationships. Although local and landscape environmental variables are generally included in studies of habitat models, riparian environmental variables, such as riparian vegetation and development, are not as commonly included. In a study conducted in river basins across the United States, Meador and Goldstein (2003) found that as the integrity of the riparian zone decreased, the condition of the fish community and water quality correspondingly decreased. They further suggested that fish community structure in streams may be better indicated by riparian conditions than by land use (Meador and Goldstein 2003). Therefore the inclusion of riparian conditions is integral to identifying key determinants of habitat quality for species and communities.

Management Implications

We observed several general trends across species, spatial scales, and statistical approaches. First, tree-based approaches exhibited higher predictive power than their linear counterparts in terms of predictive power, thereby improving our understanding of fish–habitat associations. Tree-based approaches were particularly useful, as they incorporated both linear and nonlinear relationships, in addition to interactions between environmental variables, in their models (De'ath and Fabricius 2000; De'ath 2002). As such, we advocate the comparison of statistical approaches to ultimately select the statistical approach that is best suited to the properties of the data set (e.g., Guisan and Zimmermann 2000; Sharma et al. 2008).

Second, the spatial scale at which ecological processes are operating within communities ought to be considered when developing science-based conservation and management strategies, including adaptive management strategies to conserve species over spatial and temporal scales (Cushman and Huettmann 2010). We found that fish–habitat associations vary across spatial scale, suggesting that managers should focus restoration efforts on both local and riparian habitat features to conserve fish populations (Meador and Goldstein 2003; Brind'Amour and Boisclair 2006). Further, the use of MEMs allows for quantification and visual identification of the spatial scale at which environmental conditions are acting within fish communities (Borcard et al. 2004; Brind'Amour et al. 2005). MEM analyses can improve our understanding of the scale and role of spatial processes acting on fish–habitat associations.

Third, fish densities were highest in regions with natural habitat heterogeneity. Habitat heterogeneity in the form of coverage of heterogeneous substrate types, macrophytes, and woody debris provide complex habitats that offer larger amounts of refuge from predators (MacRae and Jackson 2001; Pratt and Fox 2001). There is a higher likelihood of removal of complex habitat structure, such as macrophytes and woody debris, as lakes are developed for anthropogenic use, further reducing refuge habitat available to fishes (MacRae and Jack-

son 2001) and thus decreasing densities of fish (Sass et al. 2006; Roth et al. 2007). This finding underscores the importance of maintaining habitat integrity to sustain native fish populations and communities and can help guide adaptive management strategies to conserve native fish populations (Cushman and Huettmann 2010).

Acknowledgements

References

Anderson-Sprecher, R. 1994. Model comparisons and R². Am. Sci. **48**: 113–117.

Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. **22**(1): 42–47. doi:10.1016/j.tree.2006.09.010. PMID:17011070.

Bellehumeur, C., and Legendre, P. 1997. Aggregation of sampling units: an analytical solution to predict variance. Geogr. Anal. **29**(3): 258–266. doi:10.1111/j.1538-4632.1997.tb00961.x.

Bellehumeur, C., Legendre, P., and Marcotte, D. 1997. Variance and spatial scales in a tropical rain forest: changing the size of sampling units. Plant Ecol. **130**(1): 89–98. doi:10.1023/A:1009763830908.

Blanchet, F.G., Legendre, P., and Borcard, D. 2008. Forward selection of explanatory variables. Ecology, **89**(9): 2623–2632. doi:10.1890/07-0986.1. PMID:18831183.

Borcard, D., and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. **153**(1–2): 51–68. doi:10.1016/S0304-3800(01)00501-4.

Borcard, D., Legendre, P., Avois-Jacquet, C., and Tuomisto, H. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology, **85**(7): 1826–1832. doi:10.1890/03-3111.

Bouchard, J., and Boisclair, D. 2008. The relative importance of local, lateral, and longitudinal variables on the development of habitat quality models for a river. Can. J. Fish. Aquat. Sci. **65**(1): 61–73. doi:10.1139/f07-140.

Boyce, M.S., Vernier, P.R., Nielsen, E., and Schmiegelow, F.K.A. 2002. Evaluating resource selection functions. Ecol. Model. **157**(2–3): 281–300. doi:10.1016/S0304-3800(02)00200-4.

Breiman, L. 1996. Bagging predictors. Mach. Learn. **24**(2): 123–140. doi:10.1007/BF00058655.

Breiman, L. 2001. Random forests. Mach. Learn. **45**(1): 5–32. doi:10.1023/A:1010933404324.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984, Classification and regression trees. Wadsworth International Group, Belmont, California.

Brind'Amour, A., and Boisclair, D. 2006. Effect of the spatial arrangement of habitat patches on the development of fish habitat models in the littoral zone of a Canadian Shield lake. Can. J. Fish. Aquat. Sci. **63**(4): 737–753. doi:10.1139/f05-249.

Brind'Amour, A., Boisclair, D., Legendre, P., and Borcard, D. 2005. Multiscale spatial distribution of a littoral fish community in relation to environmental variables. Limnol. Oceanogr. **50**(2): 465–479. doi:10.4319/lo.2005.50.2.0465.

Buhlmann, P. 2004. Bagging, boosting and ensemble methods. *In* Handbook of computational statistics: concepts and methods. *Edited by* J. Gentle, W. Hardle, and Y. Mori. Springer, New York. pp. 877–907.

Cooper, S.D., Diehl, S., Kratz, K., and Sarnelle, O. 1998. Implications of scale for patterns and processes in stream ecology. Aust. J. Ecol. **23**(1): 27–40. doi:10.1111/j.1442-9993.1998.tb00703.x.

Cushman, S.A., and Huettmann, F. 2010. Spatial complexity, informatics and wildlife conservation. Springer, Tokyo.

Cutler, D.R., Edwards, T.C., Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. 2007. Random forests for classification in ecology. Ecology, **88**(11): 2783–2792. doi:10.1890/07-0539.1. PMID:18051647.

De'ath, G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. Ecology, **83**: 1105–1117.

De'ath, G. 2007. Boosted trees for ecological modelling and prediction. Ecology, **88**(1): 243–251. doi:10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2. PMID:17489472.

De'ath, G., and Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, **81**(11): 3178–3192. doi:10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

Dray, S., Legendre, P., and Peres-Neto, P. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Model. **196**(3–4): 483–493. doi:10.1016/j.ecolmodel.2006.02.015.

Drew, C.A., Wiersma, Y.F., and Huetmann, F. 2011. Predictive species and habitat modelling in landscape ecology: concepts and applications. Springer, New York.

Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M., and Rosenberg, M.S. 2002. A balanced view of scale in spatial statistical analysis. Ecography, **25**(5): 626–640. doi:10.1034/j.1600-0587.2002.250510.x.

Elith, J., and Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. **40**(1): 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S., Richardson, K., Schachetti Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M., and Zimmermann, N.E. 2006. Novel methods improve predictions of species' distributions from occurrence data. Ecography, **29**: 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**(4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x. PMID:18397250.

Ellis, N., Smith, S., and Pitcher, C. 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology, **93**(1): 156–168. doi:10.1890/11-0252.1. PMID:22486096.

Evans, J.S., Murphy, M.A., Holden, Z.A., and Cushman, S.A. 2011. Modelling species distribution and change using random forest. *In*

2106

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Predictive species and habitat modeling in landscape ecology. *Edited by* C.A. Drew, Y.F. Wiersma, and F. Huettmann. Springer, New York. pp. 139–159.

Ferrier, S., Manion, G., Elith, J., and Richardson, K. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers. Distrib. **13**(3): 252–264. doi:10.1111/j.1472-4642.2007.00341.x.

Fortin, M.-J., and Dale, M.R.T. 2005. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5): 1189–1232. doi:10.1214/aos/1013203451.

Graf, R.F., Bollmann, K., Suter, W., and Bugmann, H. 2005. The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. Landsc. Ecol. **20**(6): 703–717. doi:10.1007/s10980-005-0063-7.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**(2–3): 147–186. doi:10.1016/S0304-3800(00)00354-9.

Guisan, A., Graham, C.H., Elith, J., and Huettmann, F. 2007*a*. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. **13**(3): 332–340. doi:10.1111/j.1472-4642.2007.00342.x.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S., and Peterson, A.T. 2007*b*. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecol. Monogr. **77**(4): 615–630. doi:10.1890/06-1060.1.

Hardy, S.M., Lindgren, M., Konakanchi, H., and Huettmann, F. 2011. Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaskan waters: a first open-access ensemble model. Integr. Comp. Biol. **51**(4): 608–622. doi:10.1093/icb/icr102. PMID:21873643.

Huettmann, F., and Diamond, A.W. 2006. Large-scale effects on the spatial distribution of seabird in the Northwest Atlantic. Landsc. Ecol. **21**(7): 1089–1108. doi:10.1007/s10980-006-7246-8.

Keast, A., Harker, J., and Turnbull, D. 1978. Nearshore fish habitat utilization and species associations in Lake Opinicon (Ontario, Canada). Environ. Biol. Fishes, **3**(2): 173–184. doi:10.1007/BF00691941.

Kitchell, J.F., Johnson, M.G., Minns, C.K., Loftus, K.H., Greig, L., and Olver, C.M. 1977. Percid habitat: the river analogy. J. Fish. Res. Board Can. **34**(10): 1936–1940. doi:10.1139/f77-259.

Knudby, A., Brenning, A., and LeDrew, E. 2010. New approaches to modelling fish-habitat relationships. Ecol. Model. **221**(3): 503–511. doi:10.1016/j.ecolmodel.2009.11.008.

Lawler, J.J., White, D., Neilson, R.P., and Blaustein, A.R. 2006. Predicting climate-induced range shifts: model differences and model reliability. Glob. Change Biol. **12**(8): 1568–1584. doi:10.1111/j.1365-2486.2006.01191.x.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., and Hastie, T. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshw. Biol. **50**(12): 2034–2052. doi:10.1111/j.1365-2427.2005.01448.x.

Legendre, P., and Gallagher, E. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia (Berl.), **129**(2): 271–280. doi:10.1007/s004420100716.

Legendre, P., and Legendre, L. 1998. Numerical ecology. Elsevier Science B.V., Amsterdam.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Gurevitch, J., Hohn, M., and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography, **25**(5): 601–615. doi:10.1034/j.1600-0587.2002.250508.x.

Lele, S.R., and Keim, J.L. 2006. Weighted distributions and estimation of resource selection probability functions. Ecology, **87**(12): 3021–3028. doi:10.1890/0012-9658(2006)87[3021:WDAEOR]2.0.CO;2. PMID:17249227.

Leroux, S.J., Schmiegelow, F.K.A., Cumming, S.G., Lessard, R.B., and Nagy, J. 2007. Accounting for system dynamics in reserve design. Ecol. Appl. **17**(7): 1954–1966. doi:10.1890/06-1115.1. PMID:17974334.

MacRae, P.S.D., and Jackson, D.A. 2001. The influence of smallmouth bass (*Micropterus dolomieu*) predation and habitat complexity on the structure of littoral zone fish assemblages. Can. J. Fish. Aquat. Sci. **58**: 342–351. doi:10.1139/f00-247.

Manly, B.F.J., McDonald, T.L., Thomas, D.L., and Erickson, W.P. 2002. Resource selection by animals: statistical design and analysis for field studies. 2nd ed. Springer-Verlag. New York.

Mayo, J.S., and Jackson, D.A. 2006. Quantifying littoral vertical habitat structure and fish community associations using underwater visual census. Environ. Biol. Fishes, **75**(4): 395–407. doi:10.1007/s10641-005-5150-8.

McMahon, T.E., Terrell, J.W., and Nelson, P.C. 1984. Habitat suitability information: walleye. US Fish and Wildlife Service, FWS/OBS-82/10.56.

Meador, M.R., and Goldstein, R.M. 2003. Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition, and fish community structure. Environ. Manage. **31**(4): 504–517. doi:10.1007/s00267-002-2805-5. PMID:12677296.

Mills, E.L., Leach, J.H., Carlton, J.T., and Secor, C.L. 1993. Exotic species in the Great Lakes: A history of biotic crises and anthropogenic introductions. J. Great Lakes Res. **19**(1): 1–54. doi:10.1016/S0380-1330(93)71197-1.

Ohtani, K. 2000. Bootstrapping R^2 and adjusted R^2 in regression analysis. Econ. Model. **17**(4): 473–483. doi:10.1016/S0264-9993(99)00034-6.

Oppel, S., and Huettmann, F. 2010. Using a Random Forest model and public data to predict the distribution of prey for marine wildlife management. *In* Spatial information management in animal science. *Edited by* S. Cushman and F. Huettmann. Springer, Tokyo, Japan. pp. 151–163.

Peres-Neto, P.R., and Legendre, P. 2010. Estimating and controlling for spatial structure in the study of ecological communities. Glob. Ecol. Biogeogr. **19**(2): 174–184. doi:10.1111/j.1466-8238.2009.00506.x.

Peres-Neto, P.R., Legendre, P., Dray, S., and Borcard, D. 2006. Variance partitioning of species data matrices: estimation and comparison of fractions. Ecology, **87**(10): 2614–2625. doi:10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2. PMID:17089669.

Peters, J., Baets, B., Verhoest, N., Samson, R., Degroeve, S., Becker, P., and Huybrechts, W. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecol. Model. **207**(2-4): 304–318. doi:10.1016/j.ecolmodel.2007.05.011.

Phillips, S.J., and Dudik, M. 2008. Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, **31**(2): 161–175. doi:10.1111/j.0906-7590.2008.5203.x.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modelling of species and geographic distributions. Ecol. Model. **190**(3–4): 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Plante, M., Lowell, K., Potvin, F., Boots, B., and Fortin, M.J. 2004. Studying deer habitat on Anticosti Island, Quebec: relating animal occurrences and forest map information. Ecol. Model. **174**(4): 387–399. doi:10.1016/j.ecolmodel.2003.09.035.

Prasad, A.M., Iverson, L.R., and Liaw, A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems (N.Y.), **9**(2): 181–199. doi:10.1007/s10021-005-0054-1.

Pratt, T.C., and Fox, M.G. 2001. Biotic influences on habitat selection by young-of-year walleye (*Stizostedion vitreum)* in the demersal stage. Can. J. Fish. Aquat. Sci. **58**(6): 1058–1069. doi:10.1139/f01-054.

R Development Core Team. 2010. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Rao, C.R. 1995. A review of canonical coordinates and an alternative to correspondence analysis using hellinger distance. Qüestiió, **19**: 23–63.

Ripley, B. 2007. Rpart package [online]. R Development Core Team. Available from http://cran.r-project.org/web/packages/rpart/index.html.

Robinson, B.W., Wilson, D.S., Margosian, A.S., and Lotito, P.T. 1993. Ecological and morphological differentiation of pumpkin-seed sunfish in lakes without bluegill sunfish. Evol. Ecol. **7**(5): 451–464. doi:10.1007/BF01237641.

Rossi, J.P., and Nuutinen, V. 2004. The effect of sampling unit size on the perception of the spatial pattern of earthworm (*Lumbricus terrestris* L.) middens. Appl. Soil Ecol. **27**(2): 189–196. doi:10.1016/j.apsoil.2004.03.001.

Roth, B.M., Kaplan, I.C., Sass, G.G., Johnson, P.T., Marburg, A.E., Yannarell, A.C., Havlicek, T.D., Willis, T.V., Turner, M.G., and Carpenter, S.R. 2007. Linking terrestrial and aquatic ecosystems: The role of woody habitat in lake food webs. Ecol. Model. **203**(3–4): 439–452. doi:10.1016/j.ecolmodel.2006.12.005.

Sass, G.G., Kitchell, J.F., Carpenter, S.R., Hrabik, T.R., Marburg, A.E., and Turner, M.G. 2006. Fish community and food web responses to a whole-lake removal of coarse woody habitat. Fisheries (Bethesda, Md.), **31**(7): 321–330. doi:10.1577/1548-8446(2006)31[321:FCAFWR]2.0.CO;2.

Schaffers, A.P., Raemakers, I.P., Sy'kora, K.V., and ter Braak, C.J.F. 2008. Arthropod assemblages are best predicted by plant species composition. Ecology, **89**(3): 782–794. doi:10.1890/07-0361.1. PMID:18459341.

Scott, W.B., and Crossman, E.J. 1973. Freshwater fishes of Canada. Galt House Publications Limited, Oakville, Ont.

Segal, M.R., and Xiao, Y. 2011. Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **1**(1): 80–87. doi:10.1002/widm.12.

Sharma, S., and Jackson, D.A. 2007. Fish assemblages and environmental conditions in the lower reaches of northeastern Lake Erie tributaries. J. Great Lakes Res. **33**(1): 15–27. doi:10.3394/0380-1330(2007)33[15:FAAECI]2.0.CO;2.

Sharma, S., and Jackson, D.A. 2008. Predicting smallmouth bass (*Micropterus dolomieu*) occurrence across North America under climate change: a comparison of statistical approaches. Can. J. Fish. Aquat. Sci. **65**(3): 471–481. doi:10.1139/f07-178.

Sharma, S., Walker, S.C., and Jackson, D.A. 2008. Empirical modelling of lake water–temperature relationships: a comparison of approaches. Freshw. Biol. **53**(5): 897–911. doi:10.1111/j.1365-2427.2008.01943.x.

Storch, D., Konvicka, M., Benes, J., Martinkova, J., and Gaston, K.J. 2003. Distribution patterns in butterflies and birds of the Czech Republic: separating effects of habitat and geographical position. J. Biogeogr. **30**(8): 1195–1205. doi:10.1046/j.1365-2699.2003.00917.x.

Thompson, A.R., Petty, J.T., and Grossman, G.D. 2001. Multi-scale effects of resource patchiness on foraging behaviour and habitat use by longnose dace, *Rhinichthys cataractae.* Freshw. Biol. **46**(2): 145–160. doi:10.1046/j.1365-2427.2001.00654.x.

Appendices A, B, C, and D appear on the following pages.

2108

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Appendix A

Fig. A1. Schematic description of distance-based Moran's eigenvector maps (MEM, formally called principal coordinates of neighbour matrices; PCNM).
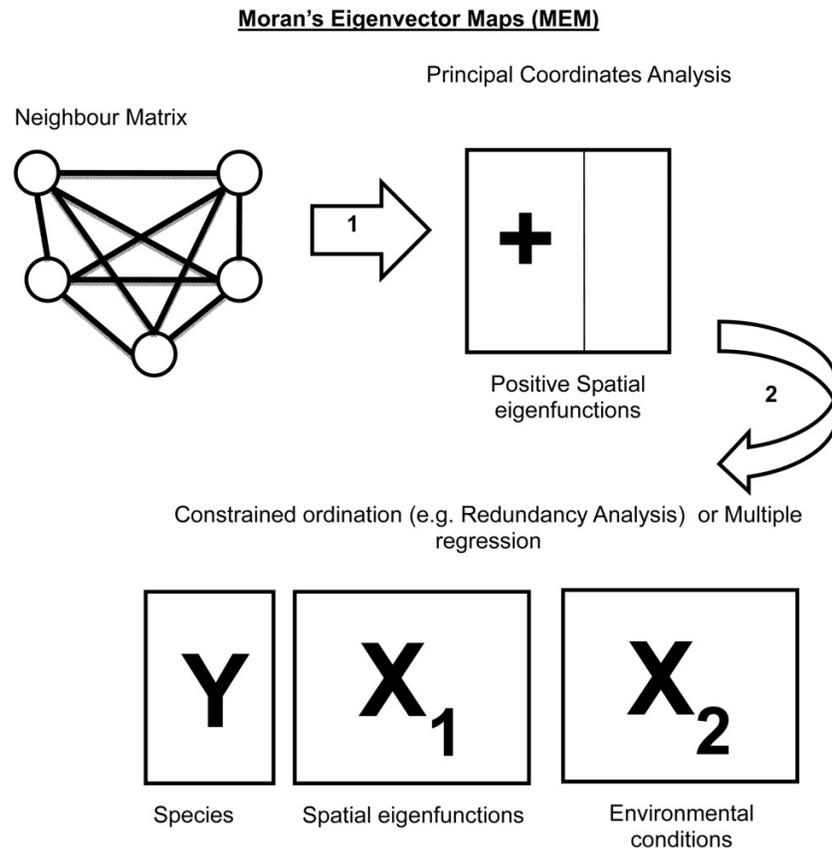
Appendix B

Table B1. Explained variation (R^2) based on regression tree and multiple regression models predicting fish density–habitat relationships in Purvis, Rond, and Violon lakes at fine and broad spatial scales.

Lake	Spatial scale	Species	% variation	
			Regression tree	Multiple regression
Purvis	S1	Yellow perch	21.56	11.25
Purvis	S1	Smallmouth bass	21.14	4.64
Purvis	S1	Pumpkinseed	20.68	3.38
Purvis	S1	Rock bass	60.61	9.51
Purvis	S5	Yellow perch	83.71	64.3
Purvis	S5	Smallmouth bass	73.72	36.28
Purvis	S5	Pumpkinseed	58.75	0
Purvis	S5	Rock bass	95.31	46.41
Rond	S1	Bluntnose minnow	53.62	19.59
Rond	S1	Banded killifish	33.22	36.81
Rond	S1	Smallmouth bass	12.84	4.97
Rond	S1	Pumpkinseed	51.61	30.52
Rond	S1	Goldfish	39.66	32.12
Rond	S5	Bluntnose minnow	84.58	82.48
Rond	S5	Banded killifish	51.92	70.81
Rond	S5	Smallmouth bass	66.35	0
Rond	S5	Pumpkinseed	89.95	67.06
Rond	S5	Goldfish	53.16	83.95
Violon	S1	Creek chub	30.21	0
Violon	S1	Fathead minnow	0	0
Violon	S1	Pearl dace	44.64	27.13
Violon	S1	Golden shiner	12.32	10.43
Violon	S1	White sucker	42.97	13.6
Violon	S1	Pumpkinseed	70.96	20.98
Violon	S1	Walleye	48.62	4.46
Violon	S5	Creek chub	31	23.58
Violon	S5	Fathead minnow	87.16	47.24
Violon	S5	Pearl dace	80.51	45.08
Violon	S5	Golden shiner	76.58	52.77
Violon	S5	White sucker	83.1	33.42
Violon	S5	Pumpkinseed	41.94	76.29
Violon	S5	Walleye	33.64	65.72

Appendix C

Table C1. Environmental variables selected for predicting fish densities in Purvis, Rond, and Violon lakes at fine (S1) and broad (S5) spatial scales using multiple linear regression (MR) and regression trees (RT).

Lake	Scale	Approach	Species	Significant predictor variables
Purvis	S1	MR	P	Cottages (+)
Purvis	S1	RT	P	Cottages (+), Depth (+), Docks (+), Pebble (–), Sand (+), Shore of slope (+), Silt (+), Trees (+)
Purvis	S5	MR	P	Cottages (+)
Purvis	S5	RT	P	Depth (+), Docks (+), Silt (+), Trees (+)
Purvis	S1	MR	R	Macrophytes (+), Trees (+)
Purvis	S1	RT	R	Boulder (–), Depth (–), Macrophytes (+), Metric boulder (+), Platforms (+), Shore of slope (+), Silt (+), Trees (+), Vegetation (–), Woody debris (+)
Purvis	S5	MR	R	Cobble (–), Pebble (+), Trees (+)
Purvis	S5	RT	R	Cottages (+), Metric boulder (–), Trees (+), Walls (+)
Purvis	S1	MR	S	Metric boulder (+)
Purvis	S1	RT	S	Bedrock (–), Cottages (+), Depth (–), Docks (–), Metric boulder (+), Trees (+), Vegetation (–), Woody debris (–)
Purvis	S5	MR	S	Metric boulder (+), Shore of slope (+)
Purvis	S5	RT	S	Bedrock (+), Cobble (–), Depth (–), Macrophytes (–), Silt (–), Trees (+)
Purvis	S1	MR	Y	Macrophytes (+)
Purvis	S1	RT	Y	Depth (–), Macrophytes (+), Trees (+)
Purvis	S5	MR	Y	Docks (–), Macrophytes (+), Metric boulder (+), Vegetation (–)
Purvis	S5	RT	Y	Depth (–), Macrophytes (+), Metric boulder (+), Shore of slope (+), Vegetation (–), Woody debris (+)
Rond	S1	MR	K	Bottom slope (+), Clay (+), Cobble (–), Metric boulder (+), Pebble (+), Trees (+), Walls (+)
Rond	S1	RT	K	Bottom slope (+), Macrophytes (+), Sand (–), Silt (–), Trees (+)
Rond	S5	MR	K	Clay (+), Pebble (+), Trees (+)
Rond	S5	RT	K	Bottom slope (+), Boulder (+), Clay (+), Silt (+)
Rond	S1	MR	L	Platforms (–), Walls (+)
Rond	S1	RT	L	Bottom slope (–), Boulder (–), Docks (+), Gravel (+), Macrophytes (+), Metric boulder (–), Silt (±)
Rond	S5	MR	L	Cottages (–), Gravel (–), Trees (+), Walls (+), Woody debris (+)
Rond	S5	RT	L	Bottom slope (–), Cottages (+), Macrophytes (+), Trees (+)
Rond	S1	MR	M	Bottom slope (+), Clay (+), Pebble (+), Metric boulder (+)
Rond	S1	RT	M	Bottom slope (+), Boulder (+), Clay (+), Docks (+), Macrophyte (–), Metric boulder (+), Sand (+), Silt (+)
Rond	S5	MR	M	Bottom slope (+), Gravel (+), Sand (–), Silt (–), Walls (–)
Rond	S5	RT	M	Bottom slope (+), Cottages (+), Metric boulder (+), Silt (+)
Rond	S1	MR	P	Bottom slope (+), Docks (+), Metric boulder (+), Pebble (+)
Rond	S1	RT	P	Bottom slope (+), Docks (+), Macrophytes (+), Metric boulder (±), Silt (+)
Rond	S5	MR	P	Metric boulder (+), Trees (–)
Rond	S5	RT	P	Docks (–), Gravel (–), Metric boulder (+)
Rond	S1	MR	S	Gravel (–)
Rond	S1	RT	S	Boulder (–), Cottages (+), Gravel (–), Macrophytes (+), Metric boulder (+), Silt (+)
Rond	S5	MR	S	Gravel (–)
Rond	S5	RT	S	Boulder (–), Sand (±), Silt (+)
Violon	S1	MR	C	Bottom slope (–), Clay (–), Pebble (–), Metric boulder (–)
Violon	S1	RT	C	Bottom slope (–), Boulder (+), Silt (+), Trees (–)
Violon	S5	MR	C	Macrophyte (–), Silt (+)
Violon	S5	RT	C	Silt (+)
Violon	S1	MR	F	Bottom slope (–), Boulder (–), Gravel (–), Macrophytes (+)
Violon	S1	RT	F	—
Violon	S5	MR	F	Boulder (–), Docks (+), Macrophytes (–)
Violon	S5	RT	F	Bedrock (+), Boulder (±), Sand (+), Silt (+)
Violon	S1	MR	G	Bottom slope (–), Metric boulder (–)
Violon	S1	RT	G	Bedrock (–), Bottom slope (±), Boulder (–)
Violon	S5	MR	G	Docks (+), Metric boulder (+), Silt (+)

Table C1 (*concluded*).

Lake	Scale	Approach	Species	Significant predictor variables
Violon	S5	RT	G	Bedrock (–), Boulder (–), Silt (–), Woody debris (+)
Violon	S1	MR	P	Boulder (–), Clay (+), Trees (+)
Violon	S1	RT	P	Boulder (±), Macrophytes (±), Trees (±), Woody debris (–)
Violon	S5	MR	P	Pebble (+), Trees (+)
Violon	S5	RT	P	Bottom slope (–), Gravel (+), Silt (–), Trees (+), Woody debris (+)
Violon	S1	MR	PD	—
Violon	S1	RT	PD	Bottom slope (–), Boulder (–), Macrophyte (+), Metric boulder (+), Silt (+), Trees (+)
Violon	S5	MR	PD	Boulder (–), Macrophytes (+)
Violon	S5	RT	PD	Bedrock (+), Boulder (–), Cobble (–), Trees (–)
Violon	S1	MR	W	Trees (+)
Violon	S1	RT	W	Boulder (+), Trees (+), Woody debris (+)
Violon	S5	MR	W	Pebble (+), Trees (+)
Violon	S5	RT	W	Bedrock (–), Metric boulder (–), Trees (–)
Violon	S1	MR	WS	Bottom slope (–), Gravel (+), Silt (+), Trees (+)
Violon	S1	RT	WS	Macrophyte (±), Metric boulder (–), Trees (+), Woody debris (+)
Violon	S5	MR	WS	Trees (+)
Violon	S5	RT	WS	Bottom slope (+), Boulder (–), Sand (+), Trees (+)

Note: +, variable has a positive effect on species density; –, variable has a negative effect on species density; ±, variable has both a positive and negative effect on species density; C, creek chub (*Semotilus atromaculatus*); F, fathead minnow (*Pimephales promelas*); G, golden shiner (*Notemigonus crysoleucas*); K, banded killifish (*Fundulus diaphanus*); L, goldfish (*Carassius auratus*); M, bluntnose minnow (*Pimephales notatus*); P, pumpkinseed (*Lepomis gibbosus*); PD, pearl dace (*Margariscus margarita*); R, rock bass (*Ambloplites rupestris*); S, smallmouth bass (*Micropterus dolomieu*); W, walleye (*Sander vitreus*); WS, white sucker (*Catostomus commersonii*); Y, yellow perch (*Perca flavescens*).

Appendix D

Table D1. Environmental variables selected for predicting densities of the fish community at fine (S1) and broad (S5) spatial scales for Purvis, Rond, and Violon lakes using redundancy analysis (RDA) and multivariate regression trees (MRT).

Lake	Spatial scale	Approach	Significant predictor variables
Purvis	S1	RDA	Bedrock, Cottages, Pebble
Purvis	S1	MRT	Bedrock, Macrophytes, Trees
Purvis	S5	RDA	Macrophytes
Purvis	S5	MRT	Boulder, Docks, Macrophytes
Rond	S1	RDA	Bottom slope, Clay, Metric boulder, Walls
Rond	S1	MRT	Boulder, Bottom slope, Metric boulder, Macrophytes, Sand
Rond	S5	RDA	Bottom slope
Rond	S5	MRT	Bottom slope, Cottages, Silt
Violon	S1	RDA	Bottom slope, Metric boulder
Violon	S1	MRT	Boulder, Bottom slope, Macrophytes, Metric boulder, Silt
Violon	S5	RDA	Docks, Macrophytes, Silt
Violon	S5	MRT	Bedrock, Silt, Trees