

Multiscale codependence analysis: an integrated approach to analyze relationships across scales

GUILLAUME GUÉNARD,^{1,3} PIERRE LEGENDRE,¹ DANIEL BOISCLAIR,¹ AND MARTIN BILODEAU²

¹Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec H3C 3J7 Canada

²Département de Mathématiques et de Statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec H3C 3J7 Canada

Abstract. The spatial and temporal organization of ecological processes and features and the scales at which they occur are central topics to landscape ecology and metapopulation dynamics, and increasingly regarded as a cornerstone paradigm for understanding ecological processes. Hence, there is need for computational approaches which allow the identification of the proper spatial or temporal scales of ecological processes and the explicit integration of that information in models. For that purpose, we propose a new method (multiscale codependence analysis, MCA) to test the statistical significance of the correlations between two variables at particular spatial or temporal scales. Validation of the method (using Monte Carlo simulations) included the study of type I error rate, under five statistical significance thresholds, and of type II error rate and statistical power. The method was found to be valid, in terms of type I error rate, and to have sufficient statistical power to be useful in practice. MCA has assumptions that are met in a wide range of circumstances. When applied to model the river habitat of juvenile Atlantic salmon, MCA revealed that variables describing substrate composition of the river bed were the most influential predictors of parr abundance at 0.4–4.1 km scales whereas mean channel depth was more influential at 200–300 m scales. When properly assessed, the spatial structuring observed in nature may be used purposefully to refine our understanding of natural processes and enhance model representativeness.

Key words: *habitat modeling; Salmo salar; scale; scale-dependent correlation; spatial model.*

INTRODUCTION

Landscape ecology and metapopulation dynamics aim at incorporating spatiotemporal information about the distribution of organisms and attributes of their habitat into ecological models (Forman and Gordon 1986, Forman 1995). In addition to its importance in obtaining dependable statistical inference tests from observational studies and field experiments, the assessment of the structures emerging from spatiotemporal organization is increasingly recognized as a cornerstone paradigm for understanding ecological processes (Wiens et al. 1993, Cottenie 2005, Wagner and Fortin 2005). Observational studies as well as large-scale experimental studies performed in the field often result in data sets whose observations are distributed across space or time. Hence, theoretical frameworks which have as an objective to analyze spatiotemporal variation in the environment benefit from quantitative methods that make use of the spatial and/or temporal structures occurring in ecological data (both responses and

explanatory variables); purposefully including these to refine models.

Methods based on sine-like eigenfunctions now exist to generate sets of orthogonal structuring variables for regularly or irregularly spaced points: spatial eigenfunctions from a connection matrix of neighboring regions or sites (Griffith 2000); PCNM, principal coordinates of neighbor matrices (Borcard and Legendre 2002); MEM, Moran's eigenvector maps (Dray et al. 2006); AEM, asymmetric eigenvector maps (Blanchet et al. 2008). These structuring variables allow researchers to dissect the spatial structure of ecological data at multiple scales (Borcard et al. 2004). When explanatory variables are involved, fractions of the total variation of a response variable can be estimated by variation partitioning analysis (Borcard et al. 2004, Peres-Neto et al. 2006). These fractions are: *a*, the proportion of variation explained uniquely with the explanatory variables; *c*, the proportion explained uniquely with the structuring variables; *b*, the proportion explained jointly by both types of variables; and *d*, the proportion not explained by any variables included in the analysis. Fraction *a* thus represents the outcome of nonspatially organized environmental processes whereas *b* + *c* represents that of both environment-driven (i.e., exogenous) and species-driven (i.e., endogenous) spatially organized processes. A method of analysis based on spatial signal is advantageous when fractions *b* and *c* account for a

Manuscript received 16 March 2009; revised 23 December 2009; accepted 28 January 2010. Corresponding Editor: H. H. Wagner.

³ Present address: Laboratoire Évolution et Diversité Biologique (EDB) UMR 5174 Université Paul-Sabatier/CNRS, 118 Route de Narbonne 4R3-b1, 31062 Toulouse Cedex 9, France. E-mail: guillaume.guenard@gmail.com

substantial proportion of the explained variation (i.e., when a strong spatial signal is observed) and environment-driven processes are suspected to occur. In this regard, it is possible to develop a computational approach that uses orthogonal structuring variables as bases to obtain scale-dependent correlation and regression coefficients between variables observed across irregular transects or surfaces. Such an approach will allow users to integrate the supplementary information associated with the positions of observations into statistical models, possibly increasing their predictive power. This modeling approach may find numerous applications whenever spatially, temporally, or phylogenetically explicit models are required.

Habitat modeling is one such domain that may benefit from the application of a spatially explicit framework (Lichstein et al. 2002, Thompson and McGarigal 2002, Schooley 2006). For instance, it has long been known that the distributions of individuals pertaining to a range of species respond to physical characteristics of their habitats (e.g., defining their ecological niche or tolerance to anthropogenic disturbance). Some of these habitat characteristics are structured in space and/or time as a consequence of their relationship with physical (e.g., geological, climatological, geomorphological, hydrological) processes that are themselves intrinsically spatially and/or temporally structured. Hence, there is interest in using the coupling of spatial/temporal signals of, for example, species abundances and habitat characteristics, instead of their discrete values at sites, for modeling purposes.

The objective of the present study is to develop an approach that will enable researchers to describe the relationship between two variables sampled at the same, irregularly spaced locations in space, or moments in time, with respect to their scales of variation. Because this technique examines the variation of two variables (response and explanatory) in relation to their spatial/temporal context respectively, it can possibly extract the information contained in the data more efficiently than traditional methods that use variables in a non-spatially/temporally explicit manner. As an exemplar scenario and in order to evaluate its performance in a concrete situation, the approach will be applied to the modeling of the river habitat of juvenile Atlantic salmon (*Salmo salar*).

METHODS

Structuring variables

Spatial or temporal series are often composed of data collected at more or less regular space or time intervals. The positions of the observations are used to calculate structuring variables. Matrix **W** contains a set of structuring variables w_i that globally (e.g., sines and cosines, or spatial eigenfunctions such as PCNMs) or locally (e.g., wavelets) describe the spatial and/or temporal relationships among the sampling units of the study. These variables are orthogonal to one another

within the set **W**, and each one has a zero sum:

$$\int_{-\infty}^{\infty} w_i(t) dt = 0 \quad \text{or} \quad \sum_{i=1}^n w_i(t) = 0 \quad (1)$$

and

$$\int_{-\infty}^{\infty} w_i(t) \times w_j(t) dt = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

or

$$\sum_{i=1}^n w_i(t) \times w_j(t) = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (2)$$

where n is the number of discrete values (points t) where functions w_i and w_j are estimated. Furthermore, the vectors in **W** are normalized to length 1, so that **W** is orthonormal. These variables are meant to describe the variation associated with the positions of observations in space or time. In the present study, we will use the form of spatial eigenfunctions called principal coordinates of neighbor matrices (PCNM; Borcard and Legendre 2002, Borcard et al. 2004, Dray et al. 2006). We will consider only PCNM eigenfunctions modeling positive autocorrelation; that is, those that have a Moran's I statistic larger than the expected value of I under H_0 . In order to comply with the property outlined in Eq. 2, the principal coordinates will be normalized to length 1, contrary to the classical PCNM functions that are usually (but not necessarily) normalized to length λ_i .

PCNM are produced by eigen-decomposition of a truncated distance matrix and are ordered by decreasing eigenvalues. In the specific case of a spatial transect or a time series with equally spaced observations, the variables have a sinusoidal aspect in terms of smoothness, continuity, and periodicity, with respective wavelengths decreasing as the rank order of the PCNM eigenfunctions increases. In that particular situation, the first PCNM variable has a wavelength nearly equal to the extent of the sampled area and the last PCNM has a wavelength nearly equal to the sampling interval. For such transect with equal sampling intervals, the approximate relationship between PCNM variable with rank i and its wavelength (λ_i) is given by

$$\lambda_i = 2 \frac{L + s}{i + 1} \quad (3)$$

where L is the extent (length) of the sampled section and s is the sampling interval. While this approximation is useful to ease the interpretation of results, it does not apply to irregular sampling designs. In the particular context of regular sampling, the wavelength is related with the size of the spatial structures prevailing in the study area or transect, and hence with the spatial scale at which these structures occur. For that reason, spatial scale and wavelength will be used interchangeably in the context of the present study. The procedure described thereafter could, however, be applied using any struc-

turing variables generated by orthogonal basis functions (eigenfunctions or other types) relevantly describing the particular phenomenon under study.

Multiscale codependence analysis (MCA): computation

Consider two random variables, a response variable (\mathbf{y}) and an explanatory variable (\mathbf{x}), each centered on its respective mean, as well as an orthonormal matrix \mathbf{W} whose columns contain a set of the structuring variables described in the previous paragraph. In the advent that the explained and explanatory variables are both correlated with a given structuring variable, the product of their simple linear correlations with the structuring variable will reflect the strength of their common correlation at the scale corresponding to that variable. Hence, a vector ($\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}}$) containing all such products of correlation coefficients (hereafter referred to as codependence coefficients) can be calculated as

$$\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}} = \frac{\mathbf{w}_i\mathbf{y}}{\sqrt{\mathbf{y}'\mathbf{y}}} \frac{\mathbf{w}_i\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}}, \quad i = 1, 2, 3, \dots, m \quad (4)$$

where m is the number of variables in matrix \mathbf{W} . This is only one possibility; another one is

$$C_{\mathbf{y},\mathbf{x};\mathbf{w}}^* = \min \left\{ \frac{(\mathbf{w}_i\mathbf{y})^2}{\mathbf{y}'\mathbf{y}}, \frac{(\mathbf{w}_i\mathbf{x})^2}{\mathbf{x}'\mathbf{x}} \right\}, \quad i = 1, 2, 3, \dots, m. \quad (5)$$

By analogy, we define the test statistic τ as the product of the Student t statistics derived from the two correlation coefficients computed with respect to a given structuring variable (\mathbf{w}_i) from a subset \mathbf{W}_s of q structuring variables in \mathbf{W} :

$$\begin{aligned} \tau_{\mathbf{y},\mathbf{x};\mathbf{w}_i \in \mathbf{W}_s} &= (n - q - 1) \frac{\mathbf{w}_i\mathbf{y}}{\sqrt{[\mathbf{y} - \mathbf{W}_s\mathbf{W}_s'\mathbf{y}][\mathbf{y} - \mathbf{W}_s\mathbf{W}_s'\mathbf{y}]}} \\ &\times \frac{\mathbf{w}_i\mathbf{x}}{\sqrt{[\mathbf{x} - \mathbf{W}_s\mathbf{W}_s'\mathbf{x}][\mathbf{x} - \mathbf{W}_s\mathbf{W}_s'\mathbf{x}]}} \quad (6) \end{aligned}$$

where n is the number of observations. In a similar manner as for Eqs. 4 and 5, another possibility would be

$$\begin{aligned} \tau_{\mathbf{y},\mathbf{x};\mathbf{w}_i \in \mathbf{W}_s} &= (n - q - 1) \min \left\{ \frac{(\mathbf{w}_i\mathbf{y})^2}{[\mathbf{y} - \mathbf{W}_s\mathbf{W}_s'\mathbf{y}][\mathbf{y} - \mathbf{W}_s\mathbf{W}_s'\mathbf{y}]}, \right. \\ &\left. \frac{(\mathbf{w}_i\mathbf{x})^2}{[\mathbf{x} - \mathbf{W}_s\mathbf{W}_s'\mathbf{x}][\mathbf{x} - \mathbf{W}_s\mathbf{W}_s'\mathbf{x}]} \right\}. \quad (7) \end{aligned}$$

It should be noted that the C^* coefficient and its associated τ^* statistic do not conserve the sign of the relationship between \mathbf{y} and \mathbf{x} with respect to \mathbf{w}_i . The τ or τ^* statistics could be tested parametrically or by independent permutations of \mathbf{y} and \mathbf{x} . Under the hypothesis that \mathbf{x} and \mathbf{y} are independent, $\mathbf{x} \sim N(0, \sigma^2I)$ and $\mathbf{y} \sim N(0, \gamma^2I)$, then τ^* is a product of two independent Student variables with $n - q - 1$ degrees of freedom. When a sufficiently large (>100) number of

degrees of freedom are available, the product distributions could be approximated using the normal product distributions (Craig 1936). Alternatively, the test could be performed using τ^* whose P value would be

$$\begin{aligned} P &= p[\min(T_1^2, T_2^2) > \tau_{\mathbf{y},\mathbf{x};\mathbf{w}_i \in \mathbf{W}_s}^*] \\ &= p[\min(F_1, F_2) > \tau_{\mathbf{y},\mathbf{x};\mathbf{w}_i \in \mathbf{W}_s}^*] = \left\{ p[F_1 > \tau_{\mathbf{y},\mathbf{x};\mathbf{w}_i \in \mathbf{W}_s}^*] \right\}^2 \quad (8) \end{aligned}$$

where T_1, T_2 are an independent Student t_{n-q-1} distribution, or F_1, F_2 are an independent Fisher $F_{1,n-q-1}$ distribution that approximate to χ_1^2 when n is very large, and p denotes cumulative probability. The sign of any $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_i}$ depends on the signs of the correlations between \mathbf{y} or \mathbf{x} and \mathbf{w}_i , being positive when both correlations are of the same sign (either positive or negative), and negative when correlations are of opposite signs. Since any eigenvector \mathbf{w}_i is equivalent to $(-1) \times \mathbf{w}_i$, the sign of the codependence coefficient only depends the sign of one of the correlations with respect to the other. A positive $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_i}$ value thus implies that both variables \mathbf{y} and \mathbf{x} follow the (spatial) trend described by \mathbf{w}_i in the same direction whereas a negative $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_i}$ value implies the variables follow this trend in opposite directions. The sign of the codependence should not be confused with the sign of the autocorrelation involved in the spatial structure.

Test procedure

Since a complete or otherwise large set of structuring variables may be available for analysis, a testing procedure is required in order to correctly select the most relevant codependence coefficients. The testing procedure developed here makes use of the orthogonality property of the structuring variables (Eq. 2). Because these variables are orthogonal to (i.e., linearly independent from) one another, the corresponding fractions of the variation they explain in the response and explanatory variables are reciprocally independent in the same fashion. This means that the calculation of a codependence coefficient for a given structuring variable, $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_i}$, does not influence the value obtained for another coefficient $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_j}$; hence the matrix-form calculation in Eq. 4. The following procedure uses the order of the coefficients (in absolute value) in vector $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}}$ as the basis for the selection procedure, which involves five steps:

- 1) Compute vector $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}}$ of the codependence coefficients (Eq. 4).
- 2) Sort the codependence coefficients in decreasing order of their absolute values.
- 3) Select the structuring variable \mathbf{w}_i associated with the highest codependence coefficient $\mathbf{C}_{\mathbf{y},\mathbf{x};\mathbf{w}_i}$ among those that have not been tested yet (i.e., not already integrated in subset \mathbf{W}_s).
- 4) Calculate the test statistic $|\tau|$ for the structuring variable \mathbf{w}_i (Eq. 6); compute its associated probability (P) using a parametric or permutation procedure.

5) Test the significance of w_i by comparing its P value to a predetermined significance level α . If w_i is significant, incorporate w_i in subset W_s and proceed again from step 3 to test another coefficient. Otherwise, stop here.

Because many candidate structuring variables may be available for testing, especially during the first steps of the testing procedure, some correction of the original, testwise, P values has to be applied in order to obtain correct, familywise, probabilities. For that purpose, a sequential version of the Šidák correction (Šidák 1967, Wright 1992) is used:

$$P' = 1 - (1 - P)^{m-q} \tag{9}$$

where P' and P are the familywise and testwise probabilities, respectively, m is the number of structuring variables available for analysis, and q is the number of structuring variables already present in subset W_s . This multiple inference correction is applied at step 4 of the aforementioned iterative testing procedure. Hence, the decision in step 5 is made using the familywise P value. The testing procedure is conducted iteratively and ends when the last statistically significant structuring variable, if any, is included in subset W_s .

Assessing goodness of fit

The codependence coefficient operates like a correlation coefficient. Its absolute value indicates the strength of the covariation of the response and the explanatory variables with a structuring variable. Assessing the goodness of fit (i.e., to what extent the values of y fitted to the model adequately represent the values observed) requires coefficients that establish such a relationship in a manner similar to a regression slope coefficient, in order to assess to what extent the response variable may be affected by the explanatory variable. For that purpose, we define the vector of coregression coefficients ($b_{y,x,w}$) as

$$b_{y,x,w_i} = \frac{w_i y}{w_i x}, \quad i = 1, 2, 3, \dots, m. \tag{10}$$

Standardized coregression coefficients ($\beta_{y,x,w}$) are similarly defined as

$$\beta_{y,x,w_i} = \sqrt{\frac{x'x w_i y}{y'y w_i x}}, \quad i = 1, 2, 3, \dots, m. \tag{11}$$

Fitted (or predicted) centered values of the response variable (\hat{y}) are obtained for a statistically significant set of coregression coefficients (b_{y,x,w_i}) by rearranging Eq. 10 and incorporating the explanatory variable (x) in the following equation:

$$\hat{y} = \sum_{i=1}^k b_{y,x,w_i} (w'_i x) w_i = \sum_{i=1}^n \{w_i w'_i\} y. \tag{12}$$

The last equality shows that the fitted or predicted values are obtained by an orthogonal projection of the

observations y onto the k -dimensional space spanned by the k selected structuring variables. Although this equality may give the impression that predictions do not use the explanatory variable x , one has to be reminded that the selection of structuring variables was done using x . Moreover, if these centered fitted values are generated using strictly exclusive subsets of structuring variables, all of which being, by definition, orthogonal from one another (i.e., Eq. 2), they will also be orthogonal to one another. The components of fitted values obtained from different explanatory variables for strictly exclusive sets of structuring variables can therefore be combined additively to provide a single, multivariate, and spatially explicit model.

More than one explanatory variable in X may show statistically significant codependence with y at a given scale (i.e., with respect to the same structuring variable). Where this situation occurs, the explanatory variable associated with the highest τ statistic (in absolute value) was selected. This selection was necessary to satisfy the aforementioned requirement for strictly exclusive sets of structuring variables; insuring that the different environmental variables that intervene in the model remain independent from one another.

Simulation study

Type I error rate.—Monte-Carlo simulations were run to estimate the type I error rate (the probability of falsely rejecting the null hypothesis, $H_0: \tau = 0$) generated by the procedure described therein when it was applied to pairs of normally distributed random y and x variables, and 10 to 1000 points were placed along a transect with regular spacing. For simplicity, each iteration of the procedure was interrupted after testing one variable, regardless of the result obtained. Because the testing was done using permutation tests, the number of permutations needed to provide dependable tests with respect to a given α threshold using reasonable computation time had to be chosen carefully. The bare minimum number of permutations (n_p) required to reach a given significance level α is

$$n_p = \alpha^{-1}. \tag{13}$$

By combining the latter equation and the reciprocal of Eq. 9, one obtains

$$n_p = \left\{ 1 - (1 - \alpha')^{1/(m-q)} \right\}^{-1} \tag{14}$$

for the number of permutations that are necessary to obtain a given, familywise, α' significance level. To insure that it remained possible to reject the null hypothesis on a familywise basis, the number of permutation used for testing in the present study was set to 10 times n_p and rounded up to the next 1000. A total of seven simulation runs were done for sample sizes of 10 to 1000 observations (Table 1).

Statistical power.—Statistical power (i.e., the probability of detecting a phenomenon of a given magnitude

TABLE 1. Conditions used in the simulations to estimate type I error rates and statistical power.

Run	Number of simulations	Sample size	Number of permutations		Minimum P value	Detection threshold (snr†)
			Type I error	Statistical power		
1	5000	10	40 999	4999	0.005	1.97:1
2	5000	25	109 999	10 999	0.005	1:1.23
3	5000	50	224 999	22 999	0.005	1:1.78
4	5000	100	454 999	44 999	0.005	1:2.53
5	4000‡	250	1 143 999	111 999	0.005	1:3.96
6	1000	500	224 999	224 999	0.05	1:4.86
7	1000	1000	448 999	448 999	0.05	1:7.40

† Signal-to-noise ratio.

‡ 1000 simulations were used for simulations on statistical power.

in the presence of noise) was investigated using Monte-Carlo simulations. Power is defined as the rate of detection of a signal in the presence of given amounts of (Gaussian) noise using a fixed significance level α ; 0.05 was used. Values of statistical power ranged from 0 to 1. For that purpose, pairs of variables (response and explanatory, y and x), each having a known amount of signal (defined as a given PCNM) and noise (i.e., random Gaussian deviates) and hence known signal-to-noise ratio (snr) were created. Again, 10 to 1000 points were placed along a transect with regular spacing. PCNM eigenfunctions were generated for these points. A six-step computational procedure was used to generate the y and x variables:

1) A structuring PCNM variable was chosen at random (e.g., w_i).

2) Two vectors of random normal deviates were generated.

3) To ensure suitable control over the amount of signal and noise in the simulated data, standardized (standard deviation = 1) residuals of the regression of these deviates were computed with respect to the selected structuring variable and used as noise components (hereafter referred to as dev_y and dev_x).

4) Signal-to-noise ratios of the response and explanatory variables (snr_y and snr_x , respectively) were obtained by the following approach using two numbers drawn at random from a uniform [0, 1] distribution (r_1 and r_2):

$$snr_y = \frac{r_1 r_2}{\sqrt{1 - r_1^2} \sqrt{1 - r_2^2}} \quad snr_x = \frac{r_1 \sqrt{1 - r_2^2}}{\sqrt{1 - r_1^2} r_2}. \quad (15)$$

5) The two pseudo-variables were constructed as follows:

$$y^* = r_1 r_2 \times w_i + \sqrt{1 - r_1^2} \sqrt{1 - r_2^2} \times dev_y$$

$$x^* = r_1 \sqrt{1 - r_2^2} \times w_i + \sqrt{1 - r_1^2} r_2 \times dev_x. \quad (16)$$

6) MCA was performed on these deviates against the complete set of structuring variables, as described in *Test procedure*. The signal was considered detected when

the test of significance of the pseudo-variables against the selected PCNM variable w_i , used to generate the pseudo-variables was statistically significant ($P' \leq 0.05$).

The signal-to-noise ratio of the relationships (snr) between the pseudo-variables (y^* and x^*) with respect to the selected structuring variable (w_i) was taken as the geometric mean of snr_y and snr_x . The above-described procedure allowed us to generate pairs of pseudo-variables with an overall mean snr of 1, with half the snr values in the range [0–1] and the other half in the range [1– ∞]. Seven simulation runs were performed for sample sizes of 10 to 1000 observations (Table 1). We used logistic regression on the results of the statistical tests (i.e., $P \leq \alpha$ coded 1 and $P > \alpha$ coded 0) to explore the relationships between statistical power and snr.

Ecological illustration: habitat of juvenile Atlantic salmon

As part of a survey to determine the factors influencing the distribution of parr (juveniles) of wild Atlantic salmon (*Salmo salar*), an endangered species, in pristine rivers, the daytime distribution of parr was surveyed by snorkeling during the summer of 2002 in the St. Marguerite River, Saguenay region, Québec, Canada (Fig. 1). A 6200 m stretch of river, ranging from site Bardsville (48°23'01.59" N, 70°12'10.05" W) to site Glasspool (48°24'01.59" N, 70°16'21.92" W), was surveyed once over a period of 10 days between 28 June and 10 August (Table 2). Water temperature (handheld thermometer, $\pm 0.5^\circ\text{C}$) was measured and the percentage of cloud cover visually estimated at the beginning and the end of each sampling day. Water discharge was estimated on the beginning of each sampling day from a fixed water level positioned at a standard location (48°16'47.25" N, 69°55'31.98" W) 50 km downstream from site Bardsville. The transect comprised a total of 310 contiguous 20-m segments. Two snorkelers swimming upstream counted the number of age I+ and II+ parr in any given 20-m segment. This surveying approach allowed an assessment of parr encounter per unit of survey effort without disturbing the upstream habitat. As the surveying effort was similar for all 20-m segments, the values of encounter per unit of effort assessed using this method represent a reliable proximate for the density of parr using the habitat (Bouchard and Boisclair 2008). Flow velocity and channel depth

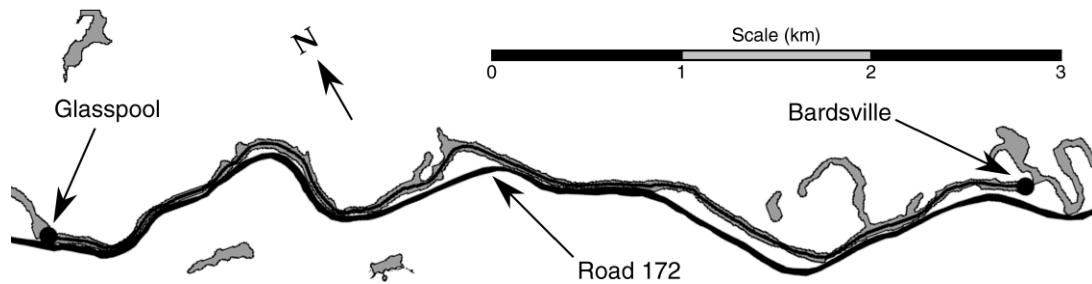


FIG. 1. Survey site: Ste-Marguerite River, Québec, Canada.

were measured in the middle of each segment using a Gurley 625A flow velocity meter (Gurley Precision Instruments, Troy, New York, USA) and a graduated pole. The measurements were made at three locations transversally and averaged: in the thalweg (i.e., the section of the water flow where the highest flow velocity is found) and at one-third and two-thirds of the distance between the thalweg and the shore that was the most distant from the thalweg. Substrate composition was estimated visually by trained observers (Latulippe and Lapointe 2001) as the percent contribution, per unit of surface, of six grain size classes. The grain size classification used in this study has been modified from Wentworth (1922) and comprises fine substrate (size < 4 mm; i.e., silt, clay, and fine to rough sand), gravel (4–32 mm), pebble (32–64 mm), cobble (64–250 mm), boulder (25–100 cm), and metric boulder (size > 1 m). Analyses were performed either on $\log_e(y + 1)$ -transformed parr abundances or using generalized linear models (GLM, quasi-Poisson family with \log_e as link function). MCA was performed after detrending (i.e., removing the linear trends from) the original data series. Since the logarithm is a monotonous positive function, logarithmic transformations do not affect the conclusions of the present study.

RESULTS

Simulation study

The type I error rates estimated from the simulation study corresponded closely to the expected error rates under the combinations of sample sizes and significance levels explored (Fig. 2; $\chi^2_{23} = 12.042$, $P = 0.97$ comparing

the observed and expected numbers of significant tests for the different thresholds). Statistical power increased with increasing sample size (Fig. 3). The minimum effect size (in terms of snr) that could be detected in 95% of the inference tests (i.e., statistical power = 0.95, Table 1) ranged from 1.97:1 to 1:7.40 over the range of sample size studied (i.e., simulated transects containing from 10 to 1000 observations). These results suggest that the testing method described herein renders correct rates of type I error in the range of conditions explored. Moreover, MCA showed good aptitude at extracting information out of noisy data, even with modest (e.g., 25–50) sample sizes.

Habitat of juvenile Atlantic salmon

Parr abundance ranged from 0 to 14 fish/20 m among 20-m segments and mean daily parr abundance ranged from 0.47 to 2.45 fish/20 m among sampling days. These differences in parr abundance were statistically significant (GLM: $F_{9,308} = 6.822$; $P = 6.30 \times 10^{-9}$) and could have been partly driven by change in one or more explanatory variables that were sampled daily (i.e., cloud cover, water temperature, and river discharge; Table 2). Therefore, an analysis of the potential effects of these variables was performed. Parr abundance was influenced positively by water temperature ($b = 0.085$; $F_{1,308} = 10.134$; parametric $P = 0.0016$) and negatively by river discharge ($b = -0.139$; $F_{1,308} = 5.8787$; parametric $P = 0.016$), but not by cloud cover ($F_{1,308} = 0.0626$; parametric $P = 0.80$). However, the influence of river discharge was no longer statistically significant when included together with water temperature in a

TABLE 2. Sampling schedule, segment sampled during each sampling day, and environmental variables estimated on a daily basis.

Date	Beginning (m)	Distance (m)	Cloud (%)	Temperature (°C)	Discharge (m ³ /s)
28 Jun	0	520	6	14.8	8.8
29 Jun	520	760	10	14.0	7.0
2 Jul	1280	1520	13	20.0	5.1
3 Jul	2800	260	45	19.0	4.7
4 Jul	3060	460	63	19.3	4.3
11 Jul	3520	480	28	14.5	3.7
19 Jul	5700	500	75	13.0	4.5
1 Aug	4000	640	30	20.8	4.0
7 Aug	4640	580	25	15.8	5.3
9 Aug	5220	480	28	17.0	3.9

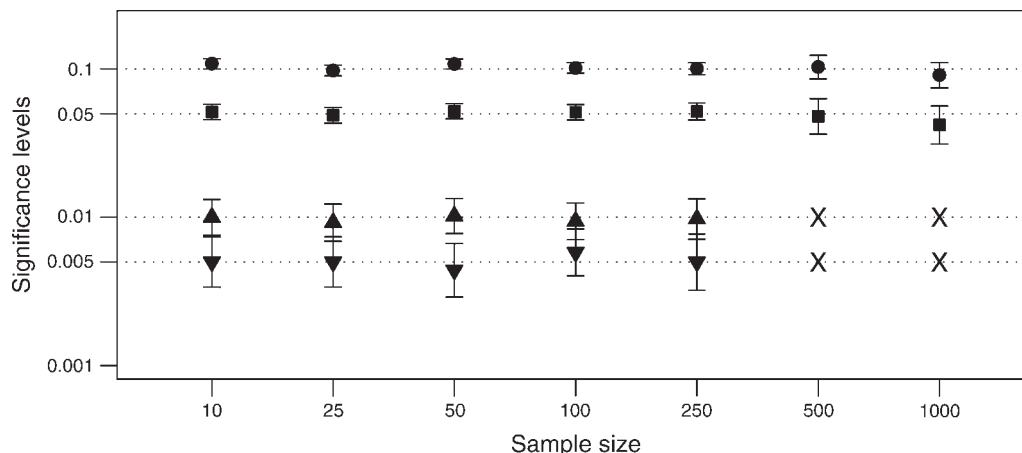


FIG. 2. Results of the simulations to assess the type I error rates generated by the codependence testing method for significance levels of 0.1 (circles), 0.05 (squares), 0.01 (up-pointing triangles), and 0.005 (down-pointing triangles) for sample sizes of 10–1000 observations. The symbol “X” represents values that cannot be reliably calculated (insufficient number of Monte-Carlo simulations). Mean values are represented with their respective 95% confidence interval bars (parametric calculations).

generalized linear model (GLM: $F_{1,307} = 1.7374$; $P = 0.19$). As a consequence, the effect of water temperature was removed by computing regression residuals from the original parr abundance data. Furthermore, parr abundance and the explanatory variables were linearly detrended with respect to transect position to remove spatial structures at scales larger than the extent of the transect. Linear detrending is suitable to obtain a more thorough identification of the spatial scales involving codependence (i.e., to prevent the spurious aliasing of structures at scales larger than the extent of sampling with smaller scale structures), but is not an absolute requirement of MCA. Eight explanatory variables were used in the subsequent analysis (Fig. 4).

A variation partitioning analysis was performed on the $\log(y + 1)$ -transformed parr abundance to provide a preliminary assessment of the importance of spatial structuring in the data set. The partitioning involved three steps. First, we computed a model describing

spatial structures in parr abundance (hereafter referred to as the spatial model) using forward-selection multiple regression with the PCNM eigenfunctions obtained from transect locations. That model involved six variables and indicated that 26.9% of the variation in parr abundance (adjusted R^2) was spatially organized. Then, we developed the environment model by computing the complete set of possible multiple regression models, using every possible combinations of explanatory variable (255 models in total), and selecting the combination associated with the highest R^2 . This model was tested for statistical significance while partialing out the variables involved in the spatial model to avoid biases that may arise from spatial autocorrelation. That approach was used because spatial autocorrelation is known to inflate the level of type I error and make the tests of significance invalid. To this end, Peres-Neto and Legendre (2010) have shown that using structuring variables as covariables in partial regression or partial

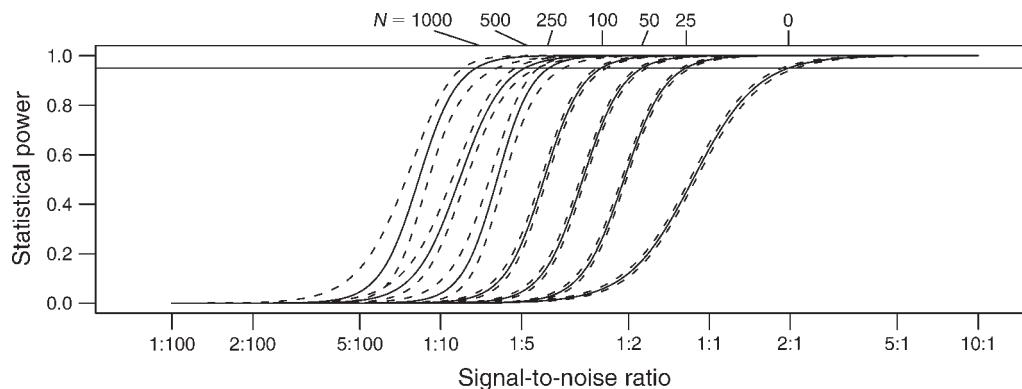


FIG. 3. Statistical power as a function of the signal-to-noise ratio (i.e., relative effect size, solid lines; note log scale) for different sample sizes (N). The dashed lines depict the 95% confidence interval of the relationships.

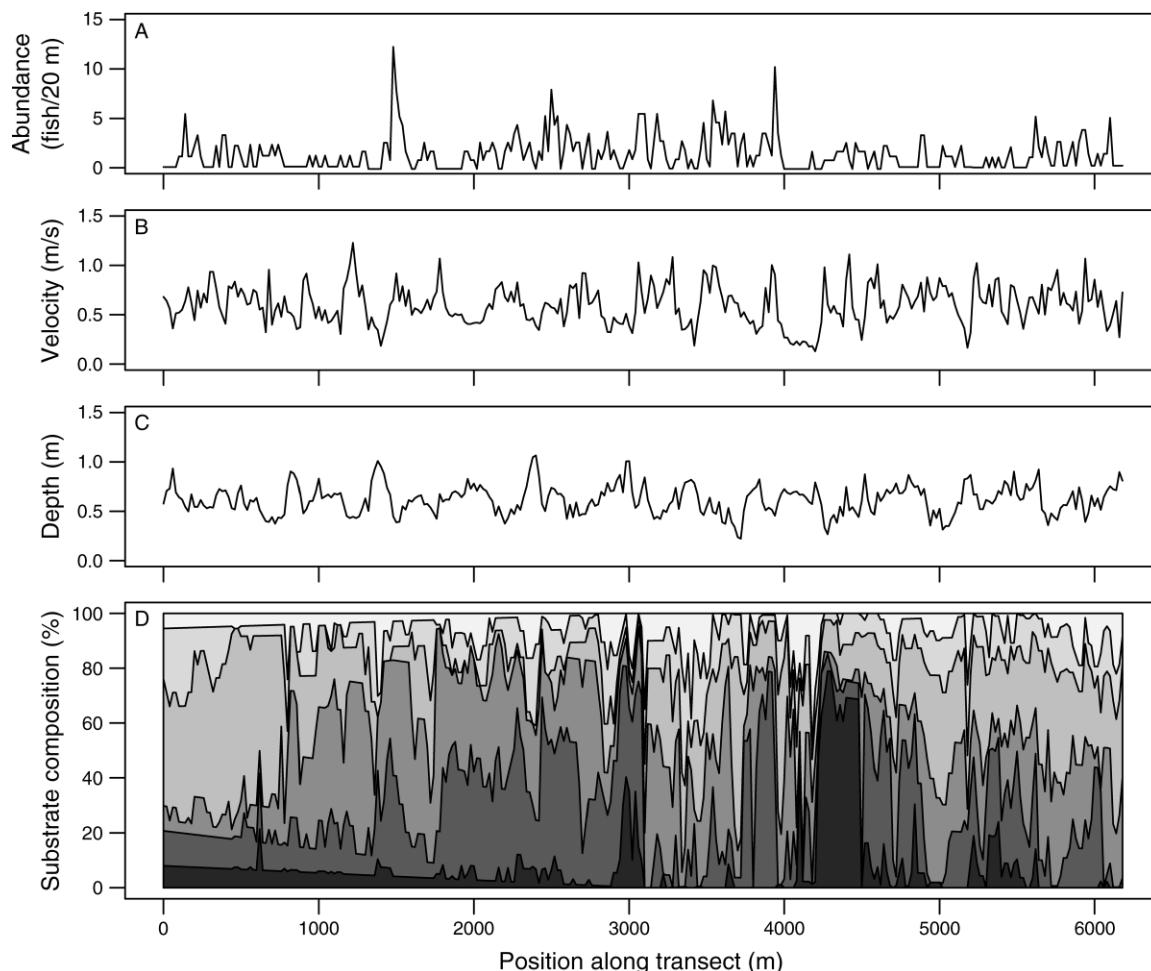


FIG. 4. (A) Abundance of Atlantic salmon parr after the effect of among-day temperature differences and linear trend were removed by regression. Also shown are (B) flow velocity, (C) channel depth, and (D) substrate composition after removal of linear trends. The six levels of shading represent the contributions of the six grain size classes of the river bed, from fine substrate (light) to metric boulder (dark).

canonical analysis is an effective way of controlling for type I error in such tests of species-environment relationships. The subset of explanatory variables that best explained parr abundance while accounting for spatial structures included three variables: channel depth, the percentage of gravel, and the percentage of boulder in the substrate. These variables explained 9.9% of the variation in parr abundance (adjusted R^2 ; $F_{3,288} = 15.17$, testwise $P < 0.0001$, familywise P over 255 tests < 0.0001). Finally, the model including both the three retained explanatory variables and the six structuring variables was computed (adjusted $R^2 = 0.248$). The unique fraction of the variation of parr abundance explained by environmental conditions was 2.0% and that explained uniquely by the PCNM eigenfunctions was 14.9% while their common fraction (b) was 7.9% (residual variation = 75.2%).

Because performing multiple tests represents an increased risk of falsely rejecting the null hypothesis,

the significance threshold of any given codependence analysis (α') had to be corrected to ensure a global, familywise, significance level (α) of 0.05. For that purpose, the Šidák approach (non-sequential version; Šidák 1967, Wright 1992) was used:

$$\alpha' = 1 - (1 - p)^{1/m}. \quad (17)$$

As a consequence, a significance level of approximately 0.00639 was used for the MCA performed on individual explanatory variables.

Thirteen combinations of explanatory variables and spatial scales explained a statistically significant fraction of the variations of parr abundance (Table 3). Seven of the eight explanatory variables and six of the 154 PCNM eigenfunctions modeling positive autocorrelation were involved in these significant relationships. At the scale of 4.1 km, parr abundance was primarily and negatively related to the percentage of pebble in the substrate. Three other variables describing substrate composi-

TABLE 3. Results of codependence analysis and associated significant statistical tests.

Variable	Scale† (m)	$C_{y,x,w}$	$b_{y,x,w}$	τ_v	v	Familywise P
Flow velocity	1131	0.036	4.17	11.67	308	0.01
Depth	265	-0.032	-6.37	-10.24‡	308	0.02
	197	-0.029	-4.39	-9.67‡	307	0.02
Fine	4147	0.052	0.10	16.95	308	<0.005
Gravel	1131	0.040	0.10	12.84	308	<0.005
	4147	0.040	0.14	13.32	307	<0.005
Pebble	4147	-0.100	-0.023	-34.36‡	308	<0.005
	2488	0.036	0.0055	14.54‡	307	<0.005
	1131	0.026	0.064	10.73	306	<0.005
	401	-0.020	-0.028	-8.61	305	0.04
Boulder	4147	0.062	0.037	20.42	308	<0.005
	401	0.027	0.021	9.18‡	307	0.02
Metric boulder	1131	-0.069	-0.027	-22.97‡	308	<0.005

Note: $C_{y,x,w}$ is the codependence coefficient, $b_{y,x,w}$ is the coregression coefficient, τ_v is the test statistic, and v is the degrees of freedom.

† Obtained using Eq. 3.

‡ These combinations of variables and spatial scales were retained in the global model and used to assess the goodness of fit.

tion—the percentages of metric boulder (negative relationship), fine substrate (positive relationship), and gravel (positive relationship)—also intervened at the scale of 4.1 km but were less closely associated with parr abundance. The percentage of pebble also influenced parr abundance at the scale of 2.5 km but this relationship was positive. A negative relationship between parr abundance and the percentage of metric boulder was observed at the scale of 1.1 km. Parr abundance at this scale was, but to a lesser extent, coupled positively with the percentages of gravel and pebble in the substrate as well as with flow velocity. The percentage of boulders was positively associated with parr abundance at the scale of 401 m while the percentage of pebble was associated negatively at that same scale. Finally, water depth was negatively related with parr abundance at the scales of 197 and 265 m.

Six combinations of environmental variables and spatial scales were used in MCA to provide a model for the assessment of the goodness of fit (Table 3; see *Methods: Assessing the goodness of fit*). Together, the percentage of pebble at scales of 4.1 and 2.5 km, the percentage of metric boulder at the scale of 1.1 km, the percentage of boulder at the scale of 400 m, and depth at the scales of 197 and 265 m explained a total of 18.9% (adjusted R^2) of the $\log(y + 1)$ -transformed variation in parr abundance (Fig. 5). For comparison, the best multiple regression model obtainable using the environmental variables explained 9.9% of the $\log(y + 1)$ -transformed variation in parr abundance using three variables. Hence, the spatialized model obtained from the MCA represents an almost two-fold improvement over this linear model. These results suggest that, by using the information associated with the distances separating the observation locations, models based on the approach described and tested in the present study can increase the amount of information that can be extracted from spatially explicit ecological data sets.

DISCUSSION

The present study demonstrates the statistical validity and the potential utility of the multiscale codependence analysis (MCA), a form of scale-specific correlation between two variables, for spatially explicit ecological data. Simulations have shown that MCA was honest in that it did not produce probabilities of falsely rejecting the null hypothesis (i.e., no codependence) that were substantially lower or larger than the expectations (i.e., the significance levels) in the range of conditions tested. Furthermore, the method was found to have the statistical power to detect relationships when they were present in the tested data set. Statistical power was found to be a function of sample size, a common situation with statistical tests based on degrees of freedom. MCA was also shown to explain a larger fraction of the variation of ecological data than methods based on linear models. The method is applicable when exogenous (environment-mediated) spatial processes are suspected to be involved as an explanation for the spatial distribution of an observed response. When such a situation occurs, both the response and explanatory variable are expected to follow similar spatial trends in the same or opposite direction, giving rise to positive or negative codependence, respectively. Data analysis can also be substantiated by combining MCA with other approaches. For instance, the presence of residual spatial variation of the response can be evidenced using multiple regression against the set of structuring variables found not to have statistically significant codependence. Furthermore, non-spatially organized environmental processes can be investigated using multiple regression against explanatory variables, partialing out known components of spatial variation.

MCA is a bivariate scale-specific correlation method. Although MCA was used in the present study to analyze the abundance data of one species, this method may be adapted to a multivariate framework. Extension to multivariate response data, such as species assemblages,

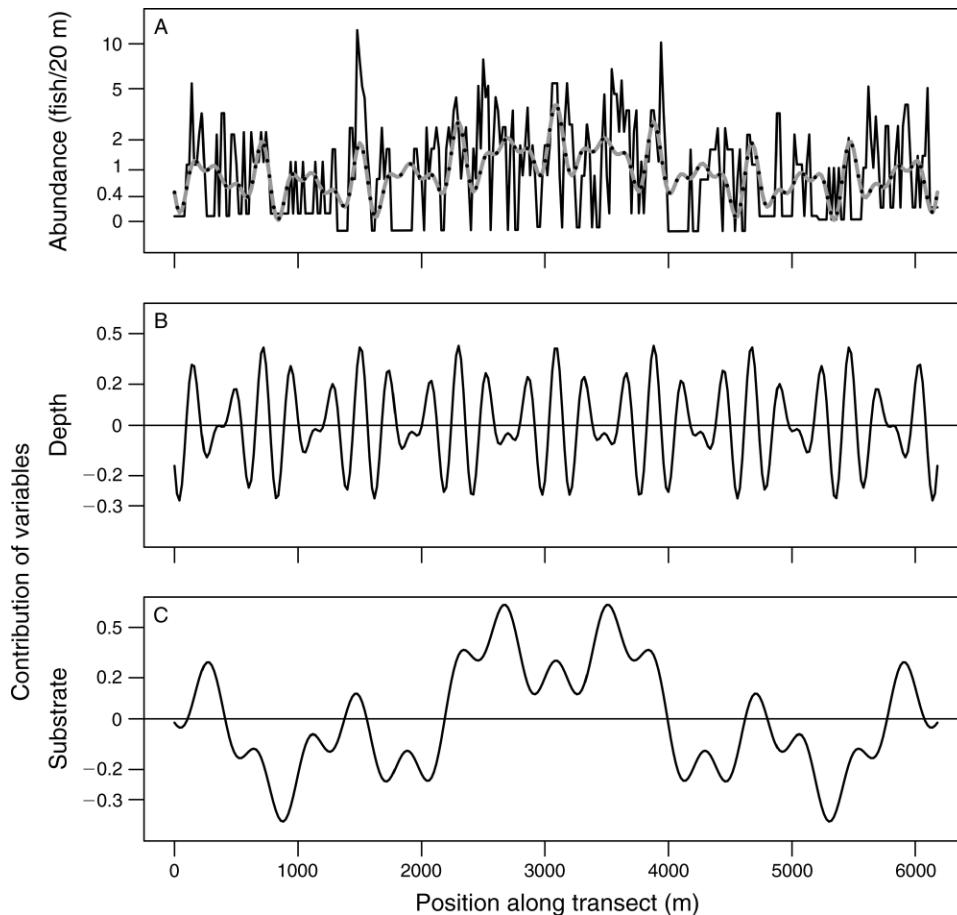


FIG. 5. (A) Observed (solid) and fitted (dotted) parr abundances, on a $\log_e(y+1)$ scale, along the river transect; (B) the fraction of parr abundance explained by channel depth; and (C) substrate composition obtained from Eq. 9. The model explains 18.9% of the $\log(x+1)$ -transformed variation in parr abundance.

may be achieved by computing MCA between the ordination axes of the community composition data obtained by principal component analysis (PCA) or correspondence analysis (CA) and the possible explanatory variables.

The computation of MCA outlined in the present study was performed using least squares fit of a common structuring variable on both a response and an explanatory variable. The testing procedure assumes that the response and explanatory variables are both the outcome of a Gaussian stochastic process (i.e., a process yielding a normally distributed error). The MCA calculated assuming normally distributed response and explanatory variables could be used for data arising from different stochastic processes by using relevant transformations. Hence, the distribution of a variable arising from a Poisson process could be approximated with the normal distribution after square-root or logarithmic transformation, while one arising from a binomial or negative-binomial processes could be approximated after arcsine or inverse hyperbolic sine

transformations, respectively (Legendre and Legendre 1998, Guan 2009). The requirement of MCA for normally distributed variables could, however, be regarded as an outcome of the computation approach used to calculate the MCA rather than an intrinsic property of MCA itself. Further work towards a generalization of MCA to other distributions, for instance using maximum-likelihood, may overcome this limitation in a similar fashion as the generalized linear model did for regression and ANOVA by providing a unifying computational framework. A second assumption to MCA is that there are linear relationships between (1) the response variable and the structuring variable, and (2) between the explanatory variable and the same structuring variable. This, in turn, implies that there is a conditional linear relationship between the response and the explanatory variable with respect to the structuring variable. A third assumption of MCA is that the estimates of both the response and the explanatory variable are reliable. The reliability of the structuring variables in representing a given spatially or

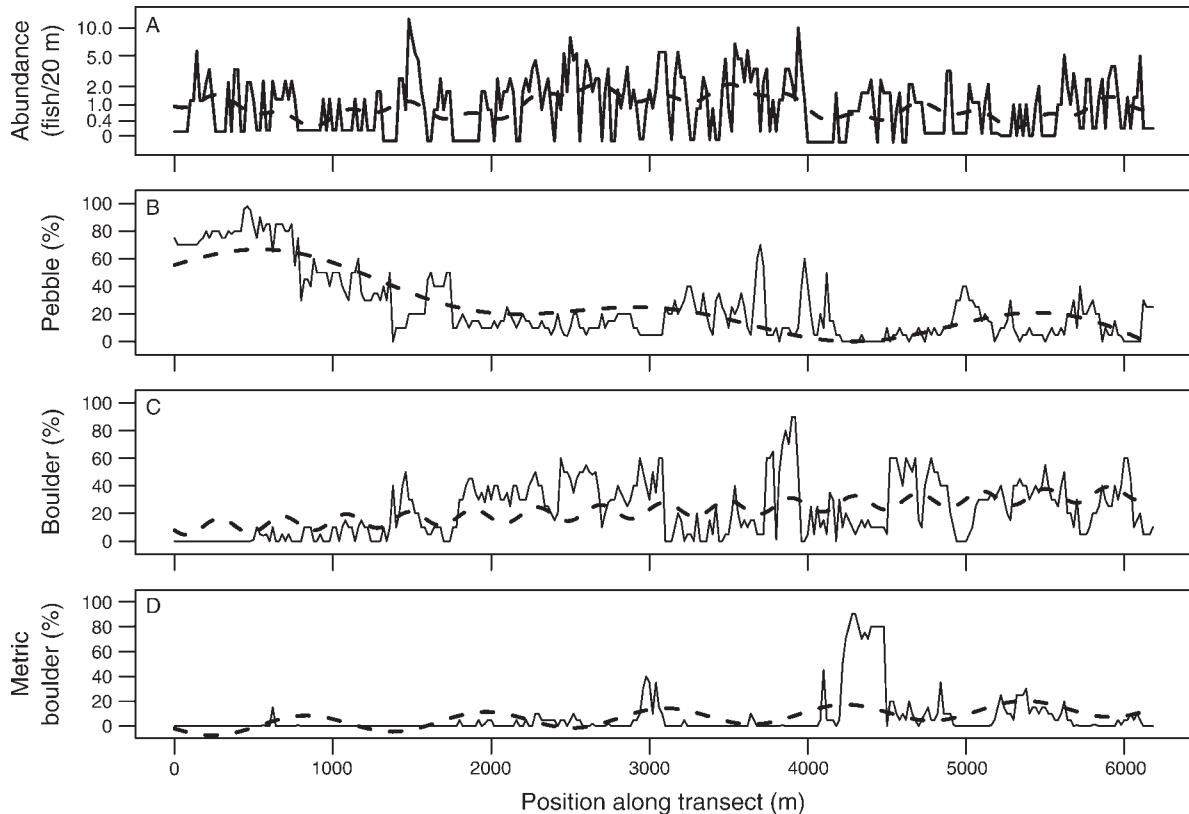


FIG. 6. (A) Observed parr abundance (solid line) and its spatially structured fraction (i.e., fitted parr abundance with respects to structuring variables of 4.1 km, 2.5 km, 1.1 km, and 0.4 km wavelength; dashed line), on a $\log_e(y + 1)$ scale. Solid lines in panels (B)–(D) present percentages of (B) pebble, (C) boulder, and (D) metric boulder; dashed lines show their spatially structured fractions: (B) scales of 4.1 and 2.5 km; (C) scale of 0.4 km; (D) scale of 1.1 km.

temporally structured process may depend on factors including, for instance, the grain size and extent of sampling, and the degree of unevenness in sampling intervals. The topic of the reliability of structuring variable is beyond the scope of the present study (see Dray et al. 2006). In the present study, PCNM, which are themselves related to Moran's I, were used to describe potential spatial structures, but any other kind of variables (e.g., wavelets) satisfying Eqs. 1 and 2 could have played the same role. Finally, the fourth assumption of MCA is the homoscedasticity of the residual variance across the fitted values, with respect to structuring variables, of the response and explanatory variables. Heteroscedasticity could be resolved using the aforementioned transformations or using a maximum likelihood approach accounting for unsteady dispersion. In MCA, the residuals of the spatial model for both the response and the explanatory variables are readily accessible, allowing easy assessments of the assumptions associated with distributions, i.e., normality and homoscedasticity, as well as residual autocorrelation.

Application of MCA to parr abundance data and environmental conditions prevailing in a river confirmed known ecological relationships among variables

and emphasized less documented interactions. For instance, parr of Atlantic salmon are generally expected to select habitats that contain a high percentage of coarse substrate (Bouchard and Boisclair 2008). However, the present study further emphasizes that boulder (25–100 cm; positive effect on parr abundance) and metric boulder (>1 m; negative influence on parr abundance) may have opposite effects on parr abundance (Table 3). In addition, the influence of environmental variables on parr abundance also varied with the spatial scale. The percent contribution of boulder to riverbed was found to have a positive effect on parr abundance at the scales of 4.1 and 0.4 km but not at smaller scales. In this transect, the percentage of boulders increased from an average of 10% (0–1300 m from the downstream limit of the transect) to reach an average of 30% (3000–4000 m from the downstream limit of the transect) and decrease to 25% (4700–5200 m from the downstream limit of the transect). The pattern observed at the scale of 4.1 km illustrates the presence in the Ste-Marguerite River of geomorphological structures defined by the downstream fining of particles (in the present case, the decrease of the percentage of boulder from ≈ 3500 m to the downstream limit of the

transect) and referred to as sedimentary links (Rice et al. 2001, Davey and Lapointe 2007). The existence of a statistically significant relationship between the abundance of parr and the percentage of boulder on the riverbed therefore supports the hypothesized role of large-scale geomorphological structures such as sedimentary links on the distribution of the biota in rivers (Rice et al. 2001; M.-È. Bédard, D. Boisclair, and M. Lapointe, *unpublished manuscript*). The pattern observed at the scale of 0.4 km may reflect the effect on fish distribution of the sequence of riffles (relatively shallow sections characterized by higher flow velocities, coarse substrate such as boulder, and high parr density) and pools (deeper river sections with lower flow velocities containing finer substrate such as sand, gravel, or pebble) observable in the Ste-Marguerite River (Davey and Lapointe 2007). The effect of sedimentary links and pool-riffle sequences on the biota may also be observed in the relationship between the abundance of parr and the percentage of pebble in the riverbed. However, the spatial variation of the percentage of pebble was explained by four structuring variables (4.1 km-, 2.5 km-, 1.1 km-, and 0.4-km wavelength; Table 3). The sign of the relationship between parr abundance and the percentage of pebble changed with respect to the spatial scale, being negative at 4.1 and 0.4 km, and positive at 2.5 and 1.1 km. The negative codependence detected between the abundance of parr and the percentage of pebble to the riverbed at the scale of 4.1 and 0.4 km mirrored the sedimentary link and the pool-riffle sequence effects mirrored the sedimentary link effect noted with boulder. The percentage of pebbles decreased from an average of 60% (0–1300 m from the downstream limit of the transect) to an average of 15% (3000–4000 m from the downstream limit of the transect) and increased to 20% (4700–5200 m from the downstream limit of the transect). The variation of the percentage of pebble followed a trend that was therefore opposite to that of parr abundance and boulder at the scale of the complete transect. In contrast, the codependence between the abundance of parr and the percentage of pebble to the riverbed was positive at the scale of 1.1 km (Table 3). The contribution of this wavelength in the distribution of the percent age of pebble to riverbed may be related to the distribution of metric boulder that has taken the form of ripples particularly in the upstream half of the transect (peak percentage of metric boulder at 3000, 4400, and 5400 m from the downstream limit of the transect; Fig. 6) via a geomorphological process that remains to be elucidated. Yet, this process may have resulted in the deposition, between zones of high percentage of metric boulder, of pebble that apparently serve as habitat for parr. This structuring suggests that the effect of pebble on parr may be contextual with a negative effect at the scale of the complete transect and pool-riffle sequences but positive when focusing on the effect of 1-km patches of pebbles located near or

between 1-km patches of metric boulder; that interaction would have been difficult to elucidate without a statistical method such as MCA. The deposition of finer particles in patches of larger substrate may explain the positive codependence between parr abundance and the percentage of gravel in the riverbed at the scale of 1.1 km, and also the positive effects of the percentages of fine substrate and gravel at the scale of 4.2 km (Table 3). Where high parr abundance was observed, boulders were often embedded in fine substrate and gravel. Finally, the negative codependence between parr abundance and water depth at scales of 200–300 m is consistent with the anticipated role of pool-riffle sequences recurrent at approximately five times the river width of the transect (full bank river width = 40–60 m). This negative codependence with water depth, together with the positive codependence with boulder and the negative codependence with pebble at a similar scale (0.4 km), therefore confirms the known preference of Atlantic salmon parr for riffles instead of pools (Bouchard and Boisclair 2008). MCA was therefore successful in refining the understanding of the effects of particular structures (opposite roles of boulder and metric boulder), in identifying unexpected processes and relationships (ripple distribution of metric blocs and its indirect role on the reversal of the effect of the percentage of pebble on parr abundance at 4.1 and 1.1 km), and at confirming the expected interactions between parrs and environmental structures at large (4.1 km; sedimentary links) and small scales (200–400 m; pool-riffle sequences).

Computational tool

An R package called “codep” is *available online*.⁴ It contains all functions needed to calculate MCA, perform the stepwise testing procedure of the codependence coefficients, and calculate fitted and residual values.

ACKNOWLEDGMENTS

We are grateful to Mariane Fradette, Marie-Ève Bédard, Sébastien Dupuis, Isabelle Barriault, and Hannah Culhane-Palmer who performed field work and to István Imre and Guillaume Bourques for relevant insights during the elaboration and application of the methods. Financial support for the field work was provided by Aquasalmo and its partners (the Fondation de la Faune du Québec, the Government of Québec-MRNF and MDEIE, the Government of Canada–Economic Development, and Alcan, Inc.). This research was supported by NSERC grants no. 7738-07 to P. Legendre, 46225-05 to D. Boisclair, and 97303-03 to M. Bilodeau.

LITERATURE CITED

- Blanchet, F. G., P. Legendre, and D. Borcard. 2008. Modelling directional spatial processes in ecological data. *Ecological Modelling* 215:325–336.
- Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153:51–68.

⁴ (<http://cran.r-project.org/src/contrib/>)

- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832.
- Bouchard, J., and D. Boisclair. 2008. The relative importance of local, lateral, and longitudinal variables on the development of habitat quality models for a river. *Canadian Journal of Fisheries and Aquatic Sciences* 65:61–73.
- Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* 8:1175–1182.
- Craig, C. C. 1936. On the frequency function $X Y$. *Annals of Mathematical Statistics* 7:1–15.
- Davey, C., and M. Lapointe. 2007. Sedimentary links and the spatial organization of Atlantic salmon (*Salmo salar*) spawning habitat in a Canadian Shield river. *Geomorphology* 83:82–96.
- Dray, S., P. Legendre, and P. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196:483–493.
- Forman, R. T. T. 1995. *Land mosaics: the ecology of landscapes and regions*. Cambridge University Press, Cambridge, UK.
- Forman, R. T. T., and M. Gordon. 1986. *Landscape ecology*. John Wiley and Sons, New York, New York, USA.
- Griffith, D. 2000. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2: 141–156.
- Guan, Y. 2009. Variance stabilizing transformations of Poisson, binomial and negative, binomial distributions. *Statistics and Probability Letters* 79:1621–1629.
- Latulippe, C., and M. Lapointe. 2001. Visual characterisation technique for gravel-cobble river bed surface sediment. *Earth Surface Processes and Landforms* 26:307–318.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Second English edition. Elsevier Science B.V., Amsterdam, The Netherlands.
- Lichstein, J. W., T. R. Simons, S. A. Shriver, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Peres-Neto, P. R., and P. Legendre. 2010. Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography* 19:174–184.
- Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87:2614–2625.
- Rice, S., M. T. Greenwood, and C. B. Joyce. 2001. Tributaries, sediment sources and the longitudinal organization of macroinvertebrate fauna along river systems. *Canadian Journal of Fisheries and Aquatic Sciences* 58:824–840.
- Schooley, R. L. 2006. Spatial heterogeneity and characteristic scales of species–habitat relationships. *BioScience* 56:533–537.
- Šidák, Z. 1967. Rectangular confidence regions for means of multivariate normal distributions. *Journal of the American Statistical Association* 62:626–633.
- Thompson, C. M., and K. McGarigal. 2002. The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York. *Landscape Ecology* 17:569–586.
- Wagner, E. H., and M.-J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology* 86:1975–1987.
- Wentworth, C. K. 1922. A scale of grade and class terms for clastic sediments. *Journal of Geology* 30:377–392.
- Wiens, J. A., N. C. Stenseth, B. Van Horne, and R. A. Ims. 1993. Ecological mechanisms and landscape ecology. *Oikos* 66:369–380.
- Wright, P. S. 1992. Adjusted p -values for simultaneous inference. *Biometrics* 48:1005–1013.