

Spatial eigenfunction analyses in stream networks: do watercourse and overland distances produce different results?

VICTOR L. LANDEIRO*, WILLIAM E. MAGNUSSON*, ADRIANO S. MELO[†], HELDER M. V. ESPÍRITO-SANTO* AND LUIS M. BINI[†]

**Programa de Pós-Graduação em Ecologia, Coordenação de Pesquisas em Ecologia, Instituto Nacional de Pesquisas da Amazônia/INPA, Manaus, AM, Brazil*

[†]Departamento de Ecologia, ICB, Universidade Federal de Goiás, Goiânia, GO, Brazil

SUMMARY

1. The use of spatial variables is a common procedure in ecological studies. The technique is based on the definition of a connectivity/distance matrix that conceptually defines the dispersal of organisms. The shortest distance between two points is a straight line. Despite the fact that a straight line may not represent the easiest dispersal path for many kinds of organisms, straight-line distances are often used to detect patterns. We argue that other types of connectivity/distance matrices will better represent dispersal paths, such as the watercourse distance for aquatic organisms (e.g. fish, shrimps).

2. We used empirical and simulated community data to evaluate the usefulness of spatial variables generated from watercourse and overland (straight-line) distances.

3. Spatial variables based on watercourse distances captured patterns that straight-line distances did not, and provided better representations of the spatial patterns generated by dispersal along a dendritic network.

Keywords: community, dispersal, overland, stream networks, watercourse

Introduction

Understanding the factors controlling the distribution of lotic organisms is one of the main objectives of stream ecology. Traditionally, stream ecologists have related the biota to environmental predictors (e.g. physicochemical measurements and substrate characteristics) to understand and predict species distribution patterns (Cummins & Lauff, 1968; Townsend & Arbuttle, 1997; Buss *et al.*, 2004; Yoshiyuki & Hajime, 2007). Recently developed statistical techniques allow us to study community distribution in relation to spatial predictors as well as environmental predictors

(Borcard & Legendre, 2002; Griffith & Peres-Neto, 2006; Peres-Neto & Legendre, 2010). These have been pivotal to the study of metacommunities (Leibold *et al.*, 2004; Holyoak, Leibold & Holt, 2005) because spatial connectivity among communities is a key aspect of metacommunity analyses. Additionally, the use of spatial predictors allows researchers to quantify, albeit indirectly, the role of dispersal in shaping beta-diversity patterns, an issue often overlooked when only environmental models are used (Bunn & Hughes, 1997).

Spatial eigenfunction analyses are frequently used to represent the variable 'space' in ecological studies, with the main objective of partitioning variance in response variables into that attributable to measured environmental variables, pure space, and shared effects of environment and space (Borcard & Legendre, 2002; Peres-Neto *et al.*, 2006; Peres-Neto &

Correspondence: Victor L. Landeiro, Instituto Nacional de Pesquisas da Amazônia INPA/CPEc, Av. Andre Araujo, 2936, CP 478 Manaus, AM CEP 69011-970, Brazil.
E-mail vllandeiro@gmail.com

Legendre, 2010). Spatial eigenfunction analysis comes under a variety of names (although they are all variations of the same theme; see Dray, Legendre & Peres-Neto, 2006), including Principal Coordinates of Neighbour Matrices (PCNM), Distance-based Eigenvectors Maps and Moran's Eigenvector Maps. However, it is not always clear what space represents, and spatial eigenvectors represent any set of variables that causes clumping in the distribution of values of the response variable(s). Space is often meant to represent dispersal limitation or some other process that is largely independent of environmental predictors, such as those proposed in neutral models of community assembly (Hubbell, 2001, 2005). However, space and environment are also highly interrelated (Tobler's first law of geography: 'Everything is related to everything else, but near things are more related than distant things'; e.g., Bjorholm *et al.*, 2008). This spatial dependency decreases our ability to identify the main processes (i.e. niche-based versus dispersal processes) and underlying patterns of community structure (Gilbert & Lechowicz, 2004). Therefore, because clumping may also result from unmeasured environmental variables, the attribution of observed patterns to dispersal processes must be performed cautiously (Diniz-Filho, Bini & Hawkins, 2003; Hawkins *et al.*, 2007).

Spatial eigenvector methods decompose the spatial variability into a set of explanatory spatial variables that represents independent propositions of how local communities are interlinked (Ramette & Tiedje, 2007). As the created variables are statistically orthogonal, they are not collinear. Standard methods to construct spatial variables (e.g. PCNM) are generally based on the use of a Euclidean distance matrix between sampling sites. This distance matrix is then submitted to a Principal Coordinate Analysis whose axes (eigenvectors) are used as spatial explanatory variables in univariate or multivariate analyses (see Borcard & Legendre, 2002 for more details about PCNM). The eigenvectors associated with high eigenvalues represent broad-scale patterns of relationships among sampling units, whereas those with low eigenvalues represent fine-scale patterns (Griffith & Peres-Neto, 2006). Euclidean distances may be appropriate for aquatic organisms that migrate over land, such as flying aquatic insects (Bilton, Freeland & Okamura, 2001). On the other hand, many organisms, such as fish and immature stages of aquatic insects,

are mostly unable to migrate in this way. For these groups, the distance between two sites might be better defined by the length of the watercourse between two points (i.e. the distance along the network pathway; Ganio, Torgersen & Gresswell, 2005; Chaput-Bardy *et al.*, 2009; Brown & Swan, 2010). In fact, several types of connectivity matrices among stream sites can be generated (Fullerton *et al.*, 2010).

Space per se cannot be considered an explanation of ecological variability (Leduc *et al.*, 1992). Thus, a significant relationship between spatial variables (eigenvectors) and raw species data tables could indicate the existence of an underlying abiotic or biotic process with a spatial component. From an ecological point of view, a set of spatial variables derived from overland distances is likely to represent a large-scale gradient in climatic conditions and other abiotic factors, whereas a set of spatial variables derived from watercourse distances is more likely to relate to dispersal limitation.

To increase our understanding of how spatial processes regulate biological communities and increase the variance explained by statistical models, new analytical methods have been proposed to incorporate other kinds of connectivity among sites in stream networks, such as autocovariance models (Peterson & Ver-Heef, 2010) and asymmetric eigenvector maps (Blanchet, Legendre & Borcard, 2008). It could be argued that the results of two analyses of variance partitioning, the first one based on spatial variables generated by using overland distances and the second by using watercourse distances (e.g. Beisner *et al.*, 2006; Nabout *et al.*, 2009), would be equivalent because of the flexibility of eigenfunction spatial analyses (Griffith & Peres-Neto, 2006). This is expected because, in both cases, several spatial variables with different spatial structures are generated. However, the equivalence of different types of connectivity matrices for spatial eigenfunction analyses has not been demonstrated.

We investigated whether spatial variables derived from watercourse distances explain more of the variance in community structure than spatial variables based on overland distances for two groups of aquatic organisms (fish and immature stages of caddisflies) in first to third order streams in a tropical forest. We also generate artificial communities, in which spatial patterns were caused only by dispersal limitation, to evaluate whether spatial variables based

on watercourse distances explain more of the variance in community structure than those based on overland (Euclidean) distances, when the model is essentially neutral (no effects of environmental variables).

Methods

Empirical field data

We used four sets of field data gathered in streams in Ducke Reserve (02°53'S, 59°58'W; near the city of Manaus, Central Amazon, Brazil). Three data sets include fish samples taken from 30 stream sites (Fig. 1) on three sampling dates (Espírito-Santo *et al.*, 2009). Six physicochemical variables were also measured: pH, conductivity, width, depth, dissolved oxygen and water velocity. Sampling details can be found in Espírito-Santo *et al.* (2009).

The fourth data set concerns caddisfly samples (Pes, 2005) taken at 27 of the same sites (Fig. 1), with three benthic sample units of 2.25 m², separated by at least 5 m, at each site. Larvae were collected using a d-net and/or a Surber sampler (both with mesh size of 250 µm) and individuals attached to bedrock or stones were removed using tweezers and spatulas. Available substrate in the sample units was assessed, collecting leaves, sand and macrophytes using d-nets, and storing in plastic bags with 80% ethyl alcohol (except stones and large woody debris). In the laboratory, the larvae were identified to morphospecies or, when possible, species. The same physicochemical variables were measured as for the fish data sets, except for dissolved oxygen.

Simulated data

We drew manually an artificial stream network where 37 sites were placed haphazardly (Fig. 2). The shape of this network is commonly observed in nature, including many branches and confluences, and has properties adequate for our objectives. The relationship between matrices generated using overland and watercourse distances among the sites was weak, and the correlation between Euclidean and watercourse distances was relatively low, as observed for streams in the Ducke Reserve network (see Peterson & Ver-
Hoef, 2010 for other ways to generate artificial stream networks). In our simulation, there was no difference between the capacities to disperse upstream or downstream.

We used a spatially explicit, individual-based (see Zurell *et al.*, 2010 for examples) simulation procedure to produce artificial community data. First, we defined a pool of 50 species for the entire network and then randomly assigned S species to each site. Each site had a fixed carrying capacity of 500 individuals that were equally distributed among the S species present at the beginning of the process. At each time step, each individual could give birth to a single offspring with a probability b and could die with a probability d . Each new individual could migrate to any site within a threshold distance, defined as the minimum distance along the network to keep all sites connected (minimum-spanning tree; Legendre & Legendre, 1998). The probability of colonising a new site was inversely proportional to the distance from the source site. The distances among

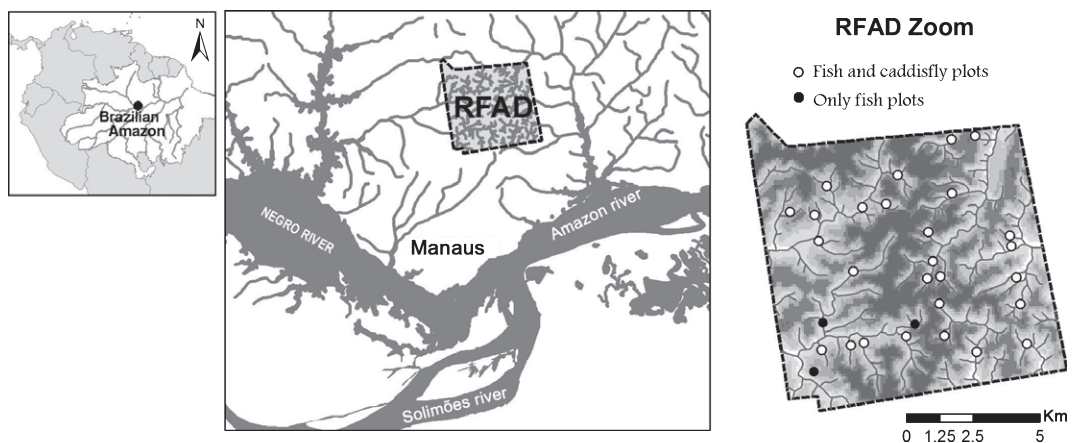


Fig. 1 Location of the Ducke Reserve (RFAD) and of the sampling plots. Fish were sampled at 30 sites and caddisfly at 27.



Fig. 2 Artificial stream network used to simulate community data. Sampling sites were haphazardly placed. Watercourse and overland distances were calculated and used in the simulation process. The probability of colonisation of a given site was inversely proportional to the distance from the source site through the watercourse corridor.

sites considered in these simulations were watercourse distances. Thus, dispersal processes were restricted to follow the network pathways. We ran the simulation for 1000 time steps. At the end of each time step and when the total number of individuals (in any given site) was larger than 500, randomly selected individuals were removed from sites until the local population was reduced to 500.

Data analysis

Data on fish and caddisfly species composition were analysed in relation to the six and five physicochemical variables, respectively, and in relation to two sets of spatial variables generated from watercourse and overland distances. All analyses were carried out in the R environment (R Development Core Team, 2009). PCNM was used to generate spatial variables, using the *pcnm* function from the *vegan* package (Oksanen *et al.*, 2010). We used partial redundancy analyses (pRDA; Borcard, Legendre & Drapeau, 1992; Legendre & Legendre, 1998) to quantify the relative importance of environmental and spatial variables in explaining the variation in community composition. For these analyses, we employed the function *varpart* from the *vegan* package. We used a stepwise selection

procedure to select spatial variables (eigenvectors) and environmental variables using the *ordistep* function from *vegan*. Community data were transformed prior to analysis using the Hellinger transformation (Legendre & Gallagher, 2001).

The threshold value used in the PCNM analysis was the minimum distance that kept all sampling sites connected using a minimum-spanning-tree procedure. However, a plateau at the centre of the Ducke Reserve separates two drainage basins. The eastern basin is connected to the streams of the western basin only by long watercourse distances, passing through the Amazon and Negro rivers (Fig. 1). This long distance may act as a barrier for dispersal of organisms from streams of one basin to streams of the other. Conventional PCNM procedure connects streams of both sides of the reserve, so we used a second truncated connectivity matrix based on the Euclidean distances, in which the western and eastern basins were 'manually' unconnected. We also used a watercourse distance matrix to generate the PCNM variables.

Hereafter, the sets of spatial variables generated by the PCNM analysis, which were based on the overland, overland with separation of the basins and watercourse distance matrices, will be referred as overland, overland-unconnected and watercourse eigenvectors, respectively. To evaluate the unique contributions of the PCNM variables generated by using these three distance matrices, we ran three partial RDA models for each of the three fish and caddisfly data sets, using environmental data and each set of eigenvectors as explanatory variables.

We analysed the simulated data using partial RDA in which spatial variables generated by watercourse and overland distances were used as two sets of explanatory variables. To quantify the shared variance explained by the two sets of spatial variables, we ran two stepwise selection procedures using the *ordistep* function to retain watercourse and overland eigenvectors to be used in the partial RDA. It is important to note that this procedure was used only to quantify the shared variance explained by the two sets of spatial variables.

Results

The correlations between overland and watercourse and between overland and overland-unconnected

distance matrices were 0.49 and 0.46, respectively. The correlation between overland-unconnected and watercourse distance was 0.62. In our artificial network, the correlation between overland and watercourse distance matrices was 0.58.

Empirical field data

For all three fish databases, analyses based on watercourse distances produced adjusted R^2 higher than those based on Euclidean (overland) distances (Table 1). RDA models applied to caddisfly data produced lower adjusted- R^2 than the RDA models applied to fish data. In general, models based on overland-unconnected distances (i.e. east and west basins unconnected) produced higher adjusted- R^2 than connected inter-basin overland distances. More spatial variables were retained in the analyses using watercourse distances, and these generally represented spatial structures at finer scales (i.e. eigenvectors with low eigenvalues).

Simulation

The high value for adjusted R^2 (0.75) obtained by RDA is an indication that the simulation procedure was effective for our objectives (i.e. we were able to generate communities with spatial patterns related to our distance matrices). There was a considerable

difference between the adjusted R^2 obtained with the spatial variables generated using overland and watercourse distances. Some spatial variables were redundant, but about 35% of the total variance explained was exclusively attributable to the spatial variables generated using watercourse distance (the distance used in the simulation). Only about 3% of the variance was exclusively attributable to the overland eigenvectors (Fig. 3).

Discussion

Stream communities are affected by processes operating at different scales, from local to regional (Heino, Louhi & Muotka, 2004; Mykrä, Heino & Muotka, 2007; Roque *et al.*, 2010). The long watercourse distance between the two Ducke Reserve basins might represent a physical barrier and constrain the dispersal of organisms. Such an effect was not detected by simple overland distances, and spatial variables generated by a watercourse distance-based eigenvector procedure explained much more of the fish community structure than the same procedure applied to overland distances.

In contrast, caddisflies have flying adults that are able to disperse over land (Collier & Smith, 1998; Wilcock *et al.*, 2007) and, in their case, the spatial variables based on overland distances accounted for more of the variability in community structure.

Table 1 Table of variance partitioning for fish and caddisfly data. Spatial variables represent spatial structures varying from broad (eigenvectors associated with high eigenvalues) to fine (low eigenvalues) scales. Spatial variables are shown in the order they were retained in the stepwise procedure (low numbers (e.g. 1, 2) represent variables with high eigenvalues). Values for each explained fraction are adjusted R^2 . Fractions are [a] pure environmental, [b] shared and [c] pure spatial. Overland-U represents the spatial variables generated by using a Euclidean distance matrix in which east and west basins are unconnected

Data	Distance	Spatial variables retained	Environmental variables retained	Fractions		
				[a]	[b]	[c]
Fish dry season 1	Overland	17, 8	Depth, oxygen dissolved, width	0.178	0.075	-0.015 ^{ns}
	Overland-U	1, 17		0.152	0.101	0.035
	Watercourse	1, 12, 22, 4		0.098	0.155	0.081
Fish rainy season	Overland	17	Depth, oxygen dissolved, pH, velocity	0.307	0.03	-0.004 ^{ns}
	Overland-U	1		0.297	0.04	0.035
	Watercourse	22, 1, 19, 4, 3, 9, 25		0.138	0.199	0.082
Fish dry season 2	Overland	17	Depth, pH	0.202	0.047	-0.017 ^{ns}
	Overland-U	1, 3, 17, 20		0.159	0.091	0.062
	Watercourse	12, 1, 22, 9, 4, 20, 19		0.077	0.172	0.088
Caddisfly	Overland	1, 3	pH, depth, conductivity	0.073	0.014	0.024
	Overland-U	1, 11		0.028	0.059	0.011 ^{ns}
	Watercourse	1, 11		0.035	0.052	0.022

^{ns}Non-significant fraction ($P > 0.05$); all other testable fractions were significant at $P \leq 0.05$.

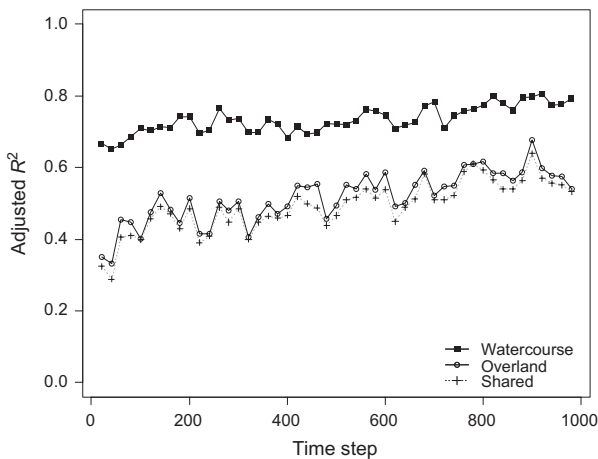


Fig. 3 Variation partitioning of simulated data. Shown are the adjusted R^2 from three separate RDA models. The first included only watercourse eigenvectors, while the second included only overland eigenvectors. The third series of values was obtained by a partial RDA that included both watercourse and overland eigenvectors to obtain the shared fraction. Note that the exclusive portion of variation explained by overland eigenvectors was minimal (c. 3%). Simulations were run for 1000 time steps and values are plotted in intervals of 20 time steps.

Moreover, the proportion of variability in caddisfly communities that was uniquely explained by spatial variables derived from watercourse distances was slightly lower than that for fishes.

Although our results showed that watercourse eigenvectors explained much more variance than overland eigenvectors, the unconnected-overland eigenvectors also explained a significant fraction of the variance in fish data. However, this explained variance was almost completely shared with watercourse eigenvectors as shown by a partial RDA using both sets of spatial predictors (variance purely attributable to unconnected-overland was 0.4% for the first dry season, <0.1% for the rainy season, and 4.9% for the second dry season). Thus, because overland and watercourse distance matrices are correlated to some extent (0.49–0.62), both distance matrices provide spatial variables that are also correlated, causing a detectable shared component. For aquatic insects that have different dispersal modes (i.e. drifting along the watercourse during immature stages and flying overland at adult stages; Poff *et al.*, 2006), the use of both types of distances might be necessary to explain species distributions. However, currently, there is no tool to tease apart the shared component, so we cannot be confident if the variation explained by one

of the two distances is spurious or not. It is important to keep this in mind when evaluating processes related to different types of distance matrices, as it is in the evaluation of gene flow by different dispersal routes (Chaput-Bardy *et al.*, 2009).

Peres-Neto & Legendre (2010) discussed the influence of the number of spatial variables on the power to detect the exclusive effect of environmental predictors in a variation-partitioning framework. Although our results also highlight this effect (i.e. a reduction in the relative contribution of the environment), they were stable in relation to the number of variables in the sense that we detected a significant environmental fraction ([a], Table 1) in most cases. Most importantly for our discussion, however, the variability of the environmental fraction was also dependent on the type of distance matrix used. Thus, our results demonstrate that the number of spatial variables and the type of distance matrix used to generate spatial variables have a profound effect on the interpretation of metacommunity models. For instance, the use of a simple overland distance matrix would suggest that a species-sorting model would be most appropriate (see Table 1), as found in several studies (e.g. Cottenie, 2005; Van der Gucht *et al.*, 2007; Vanschoenwinkel *et al.*, 2007). On the other hand, the importance of dispersal-related processes in driving metacommunities structure increased conspicuously when the watercourse distance matrix was used (Table 1; see also Beisner *et al.*, 2006 for another analysis with fish data). In these cases, the results were consistent with a mixed (species-sorting + mass effects) metacommunity, a pattern found in 29% of the 158 data sets analysed by Cottenie (2005).

Dispersal limitation appears to be the principal endogenous cause of spatial autocorrelation that is of interest to ecologists (Bahn, Krohn & O'Connor, 2008; Shurin, Cottenie & Hillebrand, 2009), and most of the discussion on the effect of 'space' is in regard to dispersal limitation. However, spatial variables generated by eigenfunction analyses do not measure this directly, and much of the variance attributed to space may be because of the effects of unmeasured environmental variables (Diniz-Filho *et al.*, 2003; Hawkins *et al.*, 2007). Also, Smith & Lundholm (2010), using simulated data, raised concerns about the use of variation partitioning as a method to tease apart the effects of niche and neutral processes, mainly because of the effects of the degree of dispersal limitation on

both pure environmental and pure spatial variance fractions. When dispersal limitation is the primary mechanism creating species distribution patterns of lotic species, as in our simulations, spatial variables based on watercourse distances do explain more of the observed variance than spatial variables based on straight-line distances (i.e. overland eigenvectors).

In the simulation study, about 60% of the variance explained was attributable to the shared component (i.e. the variability that either watercourse or overland eigenvectors was able to explain), but the exclusive variance explained by watercourse eigenvectors (note that simulated data were generated using an algorithm of dispersion only along watercourse corridors) was about 35%, while the exclusive variance explained by the overland eigenvectors was only about 3%, showing the importance of using the correct distance or connectivity matrix for eigenfunction spatial analyses.

In this study, we assumed that dispersal limitation among sites was a simple function of distance along the watercourse. However, dispersal is not necessarily as easy in an upstream direction compared to downstream (Blanchet *et al.*, 2008), or in small streams compared to large streams. If more were known about the natural history of the species, it might be possible to use more realistic distances and connectivity matrices. Even when organisms do not disperse along channel segments, other functions could represent matrix permeability and account for differences in environmental conditions that could affect dispersal. For instance, Ver-Hoef, Peterson & Theobald (2006) and Peterson, Theobald & Ver-Hoef (2007) showed that spatial models that incorporate flow direction, as well as stream distance, were more adequate than models that only use stream distance. In an application of these models, Isaak *et al.* (2010) showed that spatial models significantly outperformed their non-spatial counterparts in predicting thermal habitats of salmonids.

We conclude that the use of Euclidean distances, despite their simplicity of definition, might not be the best choice for creating spatial predictors for eigenfunction spatial analyses. This is particularly important for aquatic systems, but may well apply to terrestrial systems depending on the environmental setting (e.g. fragmented or topographically variable landscapes) and on the vagility of the taxonomic group under study.

Acknowledgments

We are grateful to Ana Pes who permitted the use of the caddisfly data set and to Jani Heino for useful discussions on the use of spatial methods in stream ecology. We also thank Thaise Emilio, who helped us with the calculation of watercourse distances, and Tadeu Siqueira who provided helpful comments on variance partitioning and spatial filter selection procedure. We thank also two anonymous reviewers for their helpful comments and Colin Townsend for careful editing and fast handling our manuscript. The fish data are available at <http://ppbio.inpa.gov.br>. VLL and HMVES received a scholarship from CNPq and are grateful to Fundação O Boticário de Proteção à Natureza and to Instituto Internacional de Educação do Brasil (IEB/Programa BECA). ASM, LMB, and WEM have been supported by several CNPq grants.

References

- Bahn V., Krohn W.B. & O'Connor R.J. (2008) Dispersal leads to spatial autocorrelation in species distributions: a simulation model. *Ecological Modelling*, **213**, 285–292.
- Beisner B.E., Peres-Neto P.R., Lindström E.S., Barnett A. & Longhi M.L. (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology*, **87**, 2985–2991.
- Bilton D.T., Freeland J.R. & Okamura B. (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.
- Bjorholm S., Svenning J.C., Skov F. & Balslev H. (2008) To what extent does Tobler's 1st law of geography apply to macroecology? A case study using American palms (Arecaceae). *BMC Ecology*, **8**, 11–20.
- Blanchet F.G., Legendre P. & Borcard D. (2008) Modelling directional spatial processes in ecological data. *Ecological Modelling*, **215**, 325–336.
- Borcard D. & Legendre P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Borcard D., Legendre P. & Drapeau P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.
- Brown B.L. & Swan C.M. (2010) Dendritic network structure constrains metacommunity properties in riverine ecosystems. *Journal of Animal Ecology*, **79**, 571–580.

- Bunn S.E. & Hughes J.M. (1997) Dispersal and recruitment in streams: evidence from genetic studies. *Journal of the North American Benthological Society*, **16**, 338–346.
- Buss D.F., Baptista D.F., Nessimian J.L. & Egler M. (2004) Substrate specificity, environmental degradation and disturbance structuring macroinvertebrate assemblages in neotropical streams. *Hydrobiologia*, **518**, 179–188.
- Chaput-Bardy A., Fleurant C., Lemaire C. & Secondi J. (2009) Modelling the effect of in-stream and overland dispersal on gene flow in river networks. *Ecological Modelling*, **220**, 3589–3598.
- Collier K.J. & Smith B.J. (1998) Dispersal of adult caddisflies (Trichoptera) into forests alongside three New Zealand streams. *Hydrobiologia*, **361**, 53–65.
- Cottenie K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175–1182.
- Cummins K.W. & Lauff G.H. (1968) The influence of substrate particle size on the microdistribution of stream macrobenthos. *Hydrobiologia*, **34**, 145–181.
- Diniz-Filho J.A.F., Bini L.M. & Hawkins B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.
- Dray S., Legendre P. & Peres-Neto P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.
- Espírito-Santo H.M.V., Magnusson W.E., Zuanon J., Mendonça F.P. & Landeiro V.L. (2009) Seasonal variation in the composition of fish assemblages in small Amazonian forest streams: evidence for predictable changes. *Freshwater Biology*, **54**, 536–548.
- Fullerton A.H., Burnett K.M., Steel E.A., Flitcroft R.L., Pess G.R., Feist B.E. *et al.* (2010) Hydrological connectivity for riverine fish: measurement challenges and research opportunities. *Freshwater Biology*, **55**, 2215–2237.
- Ganio L.M., Torgersen C.E. & Gresswell R.E. (2005) A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment*, **3**, 138–144.
- Gilbert B. & Lechowicz M.J. (2004) Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences*, **101**, 7651–7656.
- Griffith D.A. & Peres-Neto P.R. (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, 2603–2613.
- Hawkins B.A., Diniz-Filho J.A.F., Bini L.M., de Marco P. & Blackburn T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, **30**, 375–384.
- Heino J., Louhi P. & Muotka T. (2004) Identifying the scales of variability in stream macroinvertebrate abundance, functional composition and assemblage structure. *Freshwater Biology*, **49**, 1230–1239.
- Holyoak M., Leibold M.A. & Holt R.D. (2005) *Metacommunities: Spatial Dynamics and Ecological Communities*. The University of Chicago Press, Chicago.
- Hubbell S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Hubbell S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, **19**, 166–172.
- Isaak D.J., Luce C.H., Rieman B.E., Nagel D.E., Peterson E.E., Horan D.L. *et al.* (2010) Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network. *Ecological Applications*, **20**, 1350–1371.
- Leduc A., Drapeau P., Bergeron Y. & Legendre P. (1992) Study of spatial components of forest cover using partial Mantel tests and path analysis. *Journal of Vegetation Science*, **3**, 69–78.
- Legendre P. & Gallagher E.G. (2001) Ecologically meaningful transformations for ordinations of species data. *Oecologia*, **129**, 271–280.
- Legendre P. & Legendre L. (1998) *Numerical Ecology*. Elsevier, Amsterdam.
- Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.
- Mykrä H., Heino J. & Muotka T. (2007) Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. *Global Ecology and Biogeography*, **16**, 149–159.
- Nabou J.C., Siqueira T., Bini L.M. & Nogueira I.D. (2009) No evidence for environmental and spatial processes in structuring phytoplankton communities. *Acta Oecologica*, **35**, 720–726.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., O'Hara R.B. & Simpson G.L., *et al.* (2010) *Vegan: Community Ecology Package*. R package version 1.17-2. <http://CRAN.R-project.org/package=vegan> (accessed 03 January 2011)
- Peres-Neto P.R. & Legendre P. (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.
- Peres-Neto P.R., Legendre P., Dray S. & Borcard D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.

- Pes A.M.O. (2005) *Taxonomia, estrutura e riqueza das assembléias de larvas e pupas de Trichoptera (Insecta), em igarapés na Amazônia Central*. PhD Thesis. INPA/U-FAM, Manaus.
- Peterson E.E. & Ver-Hoef J.M. (2010) A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*, **91**, 644–651.
- Peterson E.E., Theobald D.M. & Ver-Hoef J.M. (2007) Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology*, **52**, 267–279.
- Poff N.L., Olden J.D., Vieira N.K.M., Finn D.S., Simmons M.P. & Kondratieff B.C. (2006) Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. *Journal of the North American Benthological Society*, **25**, 730–755.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org> (accessed 03 January 2011).
- Ramette A. & Tiedje J.M. (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2761–2766.
- Roque F.O., Siqueira T., Bini L.M., Ribeiro M.C., Tambosi L.R., Ciocheti G. *et al.* (2010) Untangling associations between chironomid taxa in Neotropical streams using local and landscape filters. *Freshwater Biology*, **55**, 847–865.
- Shurin J.B., Cottenie K. & Hillebrand H. (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia*, **159**, 151–159.
- Smith T.W. & Lundholm T. (2010) Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography*, **33**, 648–655.
- Townsend C.R. & Arbuttle C.J. (1997) The relationship between land use and physicochemistry, food resources and macroinvertebrate communities in tributaries of the Taieri River, New Zealand: a hierarchically scaled approach. *Freshwater Biology*, **37**, 177–191.
- Van der Gucht K., Cottenie K., Muylaert K., Vloemans N., Cousin S., Declerck S. *et al.* (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proceedings of the National Academy of Sciences*, **104**, 20404–20409.
- Vanschoenwinkel B., De Vries C., Seaman M. & Brendonck L. (2007) The role of metacommunity processes in shaping invertebrate rock pool communities along a dispersal gradient. *Oikos*, **116**, 1255–1266.
- Ver-Hoef J.M., Peterson E.E. & Theobald D. (2006) Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, **13**, 449–464.
- Wilcock H.R., Bruford M.W., Nichols R.A. & Hildrew A.G. (2007) Landscape, habitat characteristics and the genetic population structure of two caddisflies. *Freshwater Biology*, **52**, 1907–1929.
- Yoshiyuki T. & Hajime K. (2007) Relationship of species composition of tropical seagrass meadows to multiple physical environmental factors. *Ecological Research*, **22**, 87–96.
- Zurell D., Berger U., Cabral J.S., Jeltsch F., Meynard C.N., Munkemüller T. *et al.* (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.

(Manuscript accepted 8 December 2010)