



Spatial and species compositional networks for inferring connectivity patterns in ecological communities

Mehdi Layeghifard^{1,2}, Vladimir Makarenkov² and Pedro R. Peres-Neto^{1,*}

¹Canada Research Chair in Spatial Modelling and Biodiversity, Département des Sciences Biologiques, Université du Québec à Montréal (UQÀM), CP 8888, Succ. Centre Ville, Montréal, QC, Canada H3C 3P8,

²Département d'Informatique, Université du Québec à Montréal (UQÀM), CP 8888, Succ. Centre Ville, Montréal, QC, Canada H3C 3P8

ABSTRACT

Aim Multiple spatial and non-spatial processes are involved in determining the complex patterns underlying the spatial variation of individual species and their assemblages. This complexity, and the logistical challenges involved in following dispersal for multiple species across multiple sites, make it challenging to infer the processes underlying metacommunity spatial heterogeneity. The goal of our paper is to present a robust quantitative framework for inferring spatial patterns across multiple ecological communities.

Innovation Unlike numerous metapopulation studies that have inferred migration rates based on landscape connectivity metrics which take into account the spatial positioning of occupied and empty patches, metacommunity studies have relied on spatial predictors built without considering such information. Here, we introduce a novel method called the multi-species spatial network (MSSN) to detect and explain spatial variability in community assemblages using a graph-theoretical approach. The MSSN approach can be best described as a reconciliation between the spatial positioning of sites and their patterns of patch occupation.

Main conclusions Our simulation and real data analyses showed that our MSSN approach was better at detecting spatial patterns within metacommunities than the commonly used MEM method (Moran's eigenvector maps). In addition, our proposed framework is also useful in estimating the levels of spatial connectivity for each local community. Finally, our framework is flexible enough to incorporate different types of functions, metrics and algorithms to detect complex spatial patterns.

Keywords

Dispersal, graph theory, metacommunity, multi-species, networks, spatial variation.

*Correspondence: Pedro R. Peres-Neto, Département des Sciences Biologiques, Université du Québec à Montréal (UQÀM), CP 8888, Succ. Centre Ville, Montréal, QC, Canada H3C 3P8.
E-mail: peres-neto.pedro@uqam.ca

INTRODUCTION

Ecological entities (e.g. individuals, populations, species and communities) display complex patterns of variation in space. This spatial variation is often generated by a combination of endogenous mechanisms such as dispersal limitation (but also sociality and reproduction) and species interactions, as well as by exogenous factors such as spatially and non-spatially structured environments (e.g. the local environment, regional climate) that impose distributional patterns on species distributions via habitat filtering (Peres-Neto & Legendre, 2010; Dray *et al.*, 2012). Given that species distributions are structured by a mix of spatial and non-spatial processes and factors, the nature and origin of

spatial structures of species and their communities are not always obvious (see Leibold *et al.*, 2004, for a review). Moreover, even in cases where spatial processes are at work the complex interactions among species and communities may not necessarily leave strong spatial signatures. For example, repulsive interactions between parent trees and their seedlings can generate regular (non-spatial) patterns, or negative spatial autocorrelation (e.g. due to competitive interactions; Meyer *et al.*, 2008) and positive spatial autocorrelation (e.g. due to dispersal limitation) may actually cancel each other without leaving strong spatial signatures in species distributions (Dray, 2011).

One particular ecological level at which complex ecological interactions are evident is at the level of metacommunities

(Leibold *et al.*, 2004), i.e. spatial networks of local species assemblages connected by dispersal. Metacommunity ecology has become a framework for understanding how dispersal interacts with local community assembly to determine patterns of species distributions among patches. Metacommunity dynamics has increased our understanding of complex interactions in community ecology, mainly due to the fact that local species interactions predict much simpler patterns of community structure at larger scales than those we typically observe in natural landscapes (Huston, 1999; Holyoak *et al.*, 2005). Nevertheless the study of metacommunities is particularly challenging because of the lack of appropriate quantitative frameworks for estimating patterns of connectivity at the metacommunity level (i.e. multiple species at multiple sites; Jacobson & Peres-Neto, 2010). The main challenge is due to the fact that one cannot possibly estimate dispersal across multiple communities and multiple species directly. Moreover, dispersal dynamics can change through time, and current spatial patterns may have not been generated by recent dispersal history. Even in the case of single species distributed across patches by dispersal (i.e. metapopulations), estimating dispersal dynamics (e.g. by mark-recapture at several locations) may be technically challenging and they still may not account for the importance of past dispersal. Instead, metapopulation ecologists often infer immigration rates based on connectivity metrics that attempt to estimate the inaccessibility of a patch (or site) to potential immigrants arriving from other patches, taking into account the distribution of populations in the landscape (Hanski, 1994; Bender *et al.*, 2003). Perhaps the simplest and most common measure of patch connectivity is the distance to the nearest occupied site (e.g. Bender *et al.*, 2003). Metacommunity studies (e.g. Beisner *et al.*, 2006; Gucht *et al.*, 2007), instead, have relied on spatial predictors (e.g. geographic positioning, geographic polynomials, eigenvector maps; see Legendre *et al.*, 2005 for a review) that are quite robust for detecting spatial patterns in data but are built without considering patch connectivity inferred from information on patch occurrence of multiple species at multiple communities (i.e. a metric homologous to metapopulation connectivity in the context of metacommunities).

In order to address the challenges of assessing connectivity patterns at the metacommunity level, we introduce a novel method to detect and explain spatial variability within metacommunities using a graph-theoretic approach. A graph or network is a mathematical model of the pair-wise relations between members of a given set of objects (here sites or local communities) and has been used in a wide variety of ecological questions, including food web structure (Krause *et al.*, 2003; Luczkovich *et al.*, 2003; Banašek-Richter *et al.*, 2009), landscape connectivity (Urban & Keitt, 2001; Bodin & Norberg, 2007; Jordán *et al.*, 2007), conservation biology (Bunn *et al.*, 2000; Urban & Keitt, 2001; Rubio & Saura, 2012) and metapopulation ecology (Hanski & Ovaskainen, 2000; Brooks, 2006; Brooks *et al.*, 2008); see Urban & Keitt (2001) for an overview of the basic elements of graph theory. Here we introduce and demonstrate the robustness (based on simulations and the analysis of real data sets) and utility of a framework to investigate spatial

patterns of connectivity within metacommunities (i.e. across multiple local communities for multiple species). Our graph-based approach, hereafter referred as to multi-species spatial networks (MSSN), uses both geographic data (geographic positions of sites in the form of latitude and longitude values) and incidence data (presence-absence of species across multiple sites) as input to infer patterns of connectivity within metacommunities. As such, the proposed framework is analogous to approaches used to infer dispersal patterns in metapopulations based on patterns of patch connectivity but is more suitable for multi-species distributions.

METHODS

In graph theory, points or objects (here communities and sites are used interchangeably) in space are referred to as 'vertices' or 'nodes' and the lines (connections) linking them are called 'edges'. Therefore, a network is a collection of vertices (points) interconnected by edges (lines). The basis of our framework is to reconcile the spatial representation of the communities using a geographic network (i.e. a graph representing the spatial similarity of sites based on their spatial positioning) with the data on their species compositions (i.e. a graph representing the community similarity of sites based on their species composition). If there is a perfect match between the two (closest sites are always more similar in their species composition), then there is no need for reconciliation and the spatial graph (e.g. a dendrogram representing the geographic distribution of local communities) will perfectly represent the spatial structure in species composition across communities (i.e. metacommunity). Conversely, if there are communities that are more similar in species composition than expected by their spatial proximity, then the reconciliation between their spatial differences and species compositions can be performed by adding extra edges (links) connecting the two communities (vertices). Therefore, in our framework, the final spatial network represents the reconciliation between the spatial distribution of sites and species compositions at those sites (i.e. local communities). A diagrammatic description of the steps involved in our spatial network method is given in Fig. 1 and is based on two broad steps:

1. Build a spatial tree using pair-wise Euclidean distances between sites, computed from their geographic coordinates. Note that these geographic distances (linear) could also be transformed into other nonlinear connectivity functions such as the negative exponential, $\exp(-d)$, where d represents the distance between two sites, or other metrics of landscape resistance (see Zeller *et al.*, 2012, for a review).
2. Use the species distribution data to find extra links (reconciliations) among communities. Community similarity here was measured using the Jaccard similarity coefficient but other indexes can be certainly considered (see Legendre & Legendre, 2012, for an extensive review). Build the spatial network for the metacommunity by adding extra links to the spatial backbone tree inferred in Step 1. The technical details involved in these two steps are explained below and in Appendix S1 in Supporting Information.

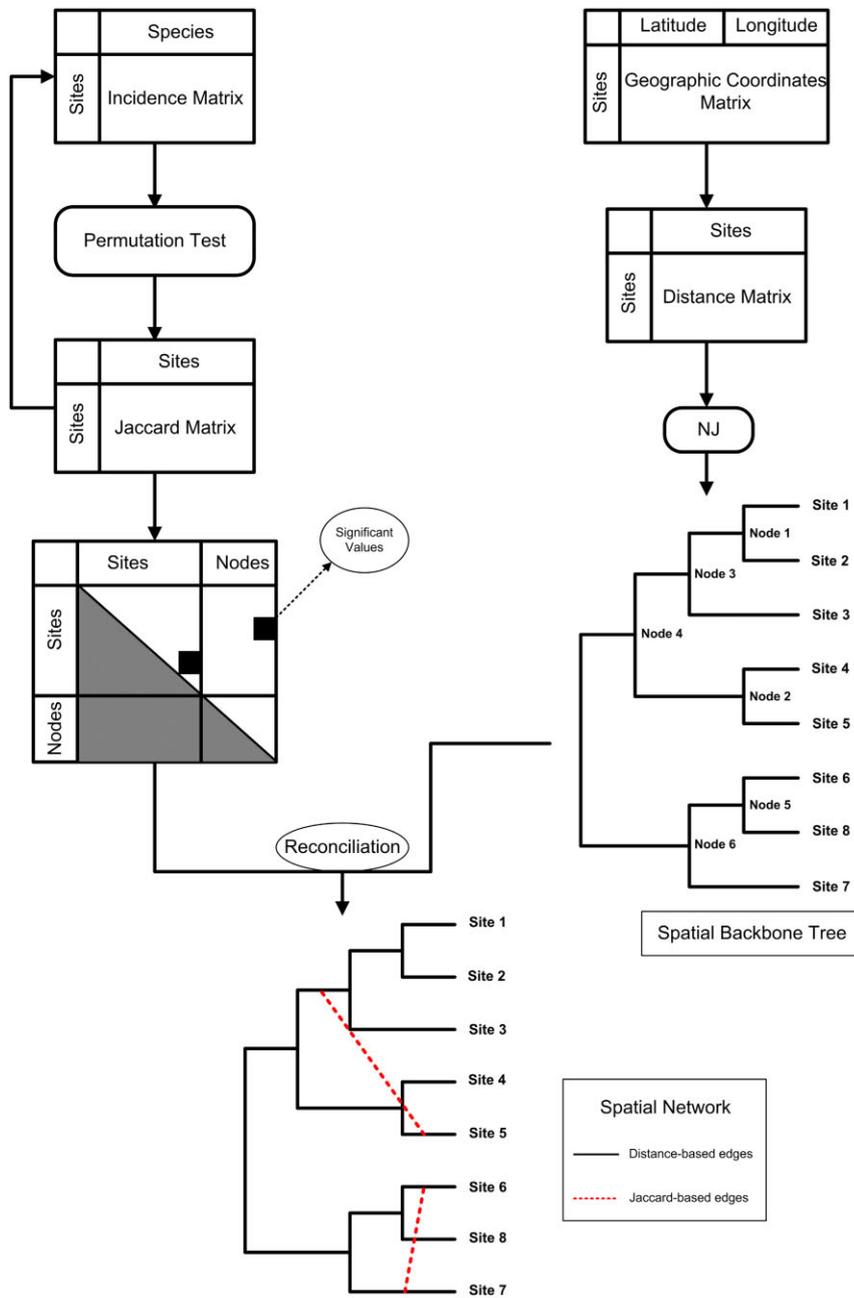


Figure 1 Diagrammatic summary of the steps involved in our spatial network method.

Step 1: building the spatial tree

Our spatial network method requires two types of input: an incidence matrix (a matrix of sites-by-species presence-absence) and a geographic positioning matrix (a matrix of sites-by-geographic coordinates). The incidence data matrix is a binary matrix of ones and zeros indicating the presence or absence of each species (columns) for each site (rows).

In order to build the spatial network, we first applied the widely used neighbour-joining method for phylogenetic reconstruction (Saitou & Nei, 1987) to build an unrooted spatial tree based on a pair-wise Euclidean geographic distance among sites (i.e. local communities). This spatial tree then serves as the

backbone of the spatial network for the local communities. A minimum length Steiner tree with 120° between all branches, which is a particular case of a phylogenetic tree, is known to generate the tree connecting all points in the plane and allows geographic information to be represented as a bifurcating minimum length tree (Cavalli-Sforza & Edwards, 1967). Neighbour-joining is well known for inferring the correct tree as long as the distance matrix is correct and ‘nearly additive’ (Felsenstein, 2004). Although in reality these conditions are rarely satisfied, neighbour-joining often constructs the correct tree topology (Mihaescu *et al.*, 2009). At any rate, different phylogeny reconstruction methods for building the spatial tree could be considered instead of neighbour-joining.

Step 2: building the multi-species spatial networks

To convert the binary spatial tree built in Step 1 into a spatial metacommunity network, we started by detecting strong (significant) connections (similarities) among local communities and then added them to the spatial tree. To do this we calculated a pair-wise species composition similarity matrix based on the Jaccard similarity coefficient (Fig. 1). Here, we only considered additional links for communities and/or nodes that shared more species than expected by chance alone. In order to estimate this probability, we randomly permuted entire rows of the incidence matrix in relation to one another and recalculated the Jaccard similarity for each permuted matrix. We repeated this step 999 times (Fig. 1) and computed a *P*-value for every pair as the number of random values smaller than or equal to their observed similarity values (+1) divided by 1000 (the addition of 1 in the numerator and the denominator represents the observed data, which is considered as a possible value of the randomized distribution). Here we considered a pair of sites as connected when both sites shared a significant number of species (here an alpha level of 0.01 was used; i.e. if the probability of rejection for any pair of local communities was equal to or greater than 0.99 we considered that they shared a greater number of species than expected by chance alone). These statistically significant links were then added to the spatial tree in order to convert it into a MSSN (i.e. our final MSSN is based on the reconciliation between the spatial representation of the communities with respect to their species compositions). Because the MSSN are also rebuilt on permuted sets of the species distribution data to estimate their significance as spatial predictors (see the section ‘Testing model with connectivity predictors’), there is no need to consider a multiple testing procedure (e.g. Bonferroni, false discovery rate) to ensure a correct Type I error when assessing the connectivity of pairs of sites in the permutation procedure above.

Building connectivity predictors for metacommunity models

In single-species metapopulation studies, the common procedure is to model species distributions (an incidence vector of presence–absence across local populations) against a predictor (or a set of predictors) of site connectivity (Hartel *et al.*, 2010; Peres-Neto & Cumming, 2010; Foltête *et al.*, 2012). In order to build a common set of connectivity predictors across local communities, we first coded our MSSN (i.e. the MSSN built in the previous section) into a weighted site-by-site matrix $\mathbf{H} = [h_{ij}]$ with each entry (h_{ij}) representing the weighted shortest path between each pair of local communities (i.e. the sum of the weights of the edges linking sites *i* and *j* with the lowest cost). Akin to metapopulation metrics (e.g. the distance to nearest occupied site), we defined weights as a compromise between the geographic proximity and community composition similarity between two communities (see Appendix S1 for a complete example with calculations). As such, these weights were

calculated based on their geographic distance weighted by their species similarities:

$$W_i = [1 - W(D)_i / \max(W(D)_i)] \times W(J)_i$$

where W_i is the overall weight of edge *i*, $W(D)_i$ is the distance-based weight of edge *i*, $\max(W(D)_i)$ is the largest Euclidean distance in the matrix so that distance values are normalized between 0 and 1 and $W(J)_i$ is the Jaccard-based weight of edge *i*; note that Jaccard distances are always bounded between 0 and 1. Here we weight community similarity in relation to geographic distance so that in cases where high levels of similarity are found between communities that are quite separate they will not be considered as connected as if the same communities were neighbouring. Note also that although here we assigned equal importance to the two measures (geographic distance and species composition similarity), differential weighting could have been certainly considered. Because both the community (Jaccard) and the spatial (geographic) weights in the above equation vary between 0 and 1, we can introduce an additional parameter ϕ that weights their relative importance as follows:

$$W_i = \phi [1 - W(D)_i / \max(W(D)_i)] \times [(1 - \phi) W(J)_i]$$

In this case, ϕ could be iterated in order to estimate the value that maximizes the predictability of assembly composition or prior knowledge (e.g. the presence of barriers) could be used to reduce the geographic over the compositional weighting, or vice versa (see Appendix S1 for more information).

In order to generate spatial predictors to be used as input for the regression models, matrix \mathbf{H} was decomposed into its eigenvectors. To perform eigenvector decomposition, we first need to double-centre matrix \mathbf{H} as follows:

$$\mathbf{H}_c = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{H} (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$$

where \mathbf{I} is an *n*-by-*n* identity matrix, $\mathbf{1}$ is an *n*-by-1 vector of ones, T denotes matrix transpose and *n* is the number of sites (local communities). We then extracted the eigenvectors from \mathbf{H}_c , which represents all orthogonal and linearly independent (i.e. orthonormality is achieved via the double-centring of the matrix \mathbf{H} into \mathbf{H}_c) spatial patterns that are possible to produce from \mathbf{H} (Griffith & Peres-Neto, 2006).

Testing the model with connectivity predictors

The final step, as in metapopulation studies (see Prugh, 2009, for a review), was to model species distributions on the basis of our connectivity predictors. Because we have multiple species, we used redundancy analysis (RDA), a regression modelling technique that can accommodate multiple response variables (species) and is the method of choice for multi-species community analyses (Legendre & Legendre, 2012). Model fit was assessed via an adjusted coefficient of determination (R^2_{adj} ; Peres-Neto *et al.*, 2006).

In order to estimate the significance of RDA when using the MSSN eigenvector predictors, we applied the following steps: (1)

calculate the observed R^2_{adj} ; (2) permute entire rows of the species matrix and recalculate the Jaccard similarity matrix and associated probability for each pair of sites (local communities); (3) based on the permuted similarity matrix, rebuild the MSSN network – note that the spatial tree was kept unchanged when recalculating the network based on the permuted similarities; (4) rebuild spatial predictors (\mathbf{H} and its eigenvectors) based on the MSSN for the permuted set; (5) calculate the R^2_{adj} of the RDA model based on the MSSN eigenvectors in Step 4; (6) repeat Step 2 to 5 999 times; (7) contrast the observed R^2_{adj} against the distribution of permuted values obtained in Step 6. The probability that a model was significant was estimated as follows: (number of random R^2_{adj} equal to or larger than observed $R^2_{\text{adj}} + 1$)/(999 + 1). The addition of 1 to the numerator and the denominator represents the observed data, which is considered as a possible value of the randomized distribution. Model significance was judged on the basis of an alpha value of 0.05. In initial simulations we noted that not rebuilding \mathbf{H} for each permuted species matrix led to inflated Type I error rates.

Assessing the performance of MSSN via simulations

Here we compared the performance (Type I error rates, statistical power and model fit, i.e. R^2_{adj}) of our approach with the most commonly used approach to model the spatial component of multi-species distributions, namely Moran's eigenvector maps (MEM; Dray *et al.*, 2006, 2012; Legendre & Legendre, 2012). MEM are the eigenvectors of a spatial matrix that simply considers the geographic proximity of sites; this differs from our MSSN approach in that spatial distributional characteristics of species are not considered. In MEM sets of orthonormal functions (i.e. eigenfunctions) are obtained from the locations of observations within a spatial framework by the eigenvalue decomposition of a spatial weighting matrix (see Griffith & Peres-Neto, 2006, for details of the calculation).

In order to contrast the performance of the MSSN method against MEM, we used both simulated and real data (see the section 'Assessing the performance of MSSN on real datasets' for an application with real data). We simulated metacommunities consisting of 2500 local communities (sites) and 50 species distributed across a squared lattice (50×50). The first step was to calculate a pair-wise geographic Euclidean distance matrix $\mathbf{D} = [d_{ij}]$ among all the 2500 communities in the landscape. Next, in order to generate spatial patterns into the species distributions within the metacommunity (lattice), we created a spatial matrix \mathbf{W} as follows:

$$\mathbf{W} = [w_{ij}] = \begin{cases} \frac{3d_{ij}}{2a} - 0.5 \left(\frac{d_{ij}}{a} \right)^3 & \text{if } d_{ij} \leq a \\ 0 & \text{if } d_{ij} > a \end{cases}$$

where a represents the range parameter (greater values represent greater autocorrelation, i.e. more spatially structured metacommunities). Next, the Cholesky decomposition was applied to \mathbf{W} . By post-multiplying the upper-triangular from the decomposed matrix by a random normally distributed

vector $N(0,1)$ with 2500 observations, we created a normally distributed variable \mathbf{X} according to a spherical variogram with a specific given range a . Because we wanted to simulate species having different levels of similarity in their distributions across local communities, we created a vector $\mathbf{b} = [b_i]$ with 50 entries varying in increments of one from $-S/2$ to $+S/2$, where S is the number of species. For each species, we created a vector of probabilities \mathbf{P}_i corresponding to the chance that the i th species occupies the j th local community according to the simulated spatial gradient \mathbf{X} as follows:

$$\mathbf{P}_i = \frac{1}{1 + \exp(-b_0 + b_i \mathbf{X} + \kappa)}$$

where $-b_0$ is a randomly generated number from a uniform distribution that changes for each species i and κ is a random normally distributed vector $N(0,1)$ with 2500 observations that introduces further noise to species distributions (i.e. so that species having very similar b -values do not end up with extremely similar probabilities). By combining all \mathbf{P}_i vectors and converting them to a binary matrix using binomial trials, we created an incidence matrix for the given metacommunity. Note that although the approach used here to simulate metacommunities does not simulate dispersal per se, our simulation protocol would have led to parallel results if we had actually simulated movement across the landscape instead. This is because in our simulation species tracked ('dispersed towards') environmental features that are themselves spatialized.

In order to contrast the power of MEM against our MSSN, \mathbf{X} was generated by considering spatial ranges (a) from 1 to 30. For Type I error estimates, \mathbf{X} was a non-spatial variable $N(0,1)$. For each range and \mathbf{X} , we simulated 500 different metacommunities that were then used to infer spatial variability using our MSSN method. Before doing so, however, some non-spatial patterns were introduced to each simulated metacommunity by replacing 0, 10 or 20% of local communities with randomly chosen communities within the metacommunity to see how the method behaves. Next, from each simulated metacommunity (30 ranges \times 500 metacommunities \times 3 (0, 10 or 20% replacement) = 45,000 metacommunities), samples containing 100 and 150 local communities, respectively, were randomly selected within each metacommunity. Samples were taken because in realistic situations we only estimate patterns of species distributions in a much smaller number of local communities in contrast to the whole metacommunity. The sampled data from the metacommunity represented an incidence matrix of species occurrences across sampled local communities as well as their matrix of geographic coordinates. These two matrices were then used as input to RDA_{MEM} and our RDA_{MSSN} method.

Assessing the performance of MSSN on real datasets

Here we used a dataset on lake fish communities across the province of Ontario, Canada. This data set, which was obtained from the Ontario Fish Distribution Database (OFDD) maintained by the Ontario Ministry of Natural Resources (OMNR),

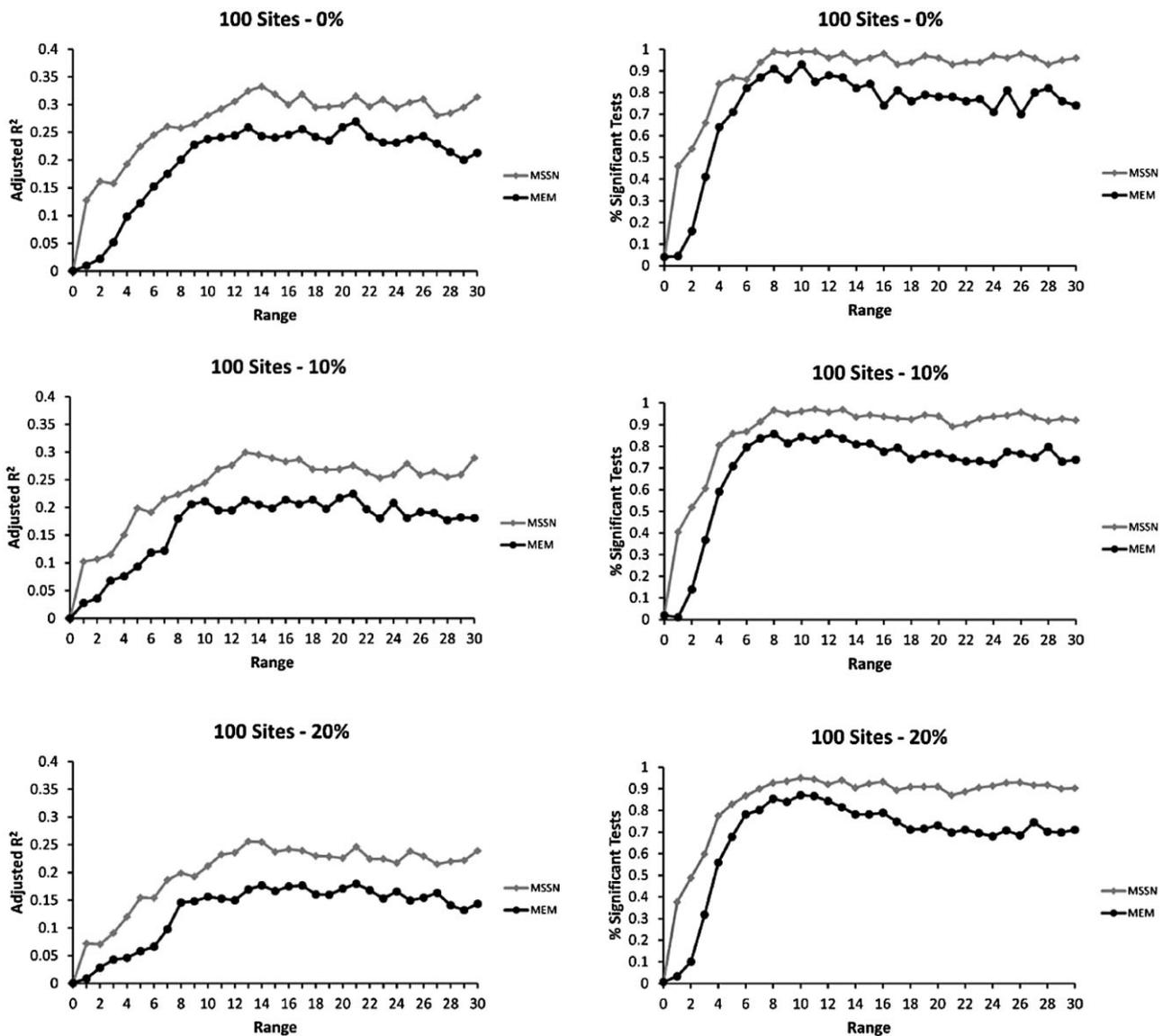


Figure 2 Contrast in performance between the multi-species spatial network (MSSN) versus Moran's eigenvector maps (MEM) on the basis of metacommunities based on 100 simulated communities. Left panels: average adjusted R^2 values for simulated landscapes with 0, 10 and 20% changes by our MSSN method and the MEM method. Right panels: Type I error (range 0) and power (range from 1 to 30) measured as the proportion of rejections ($\alpha = 0.05$) out of 500 tests for both MSSN and MEM methods.

contains the presence–absence records of 134 fish species distributed among approximately 9900 lakes as well as the geographic positions of the lakes. Given the large number of lakes, there is no available information on stream connectivity among most lakes and we instead used overland (Euclidean) distances between lakes in all cases. Note, however, that our procedure can take any type of geographic or spatial information (e.g. Euclidean distance, stream length linking lakes, functional connectivity, resistance surfaces). We used presence–absence records collected in summers between 1968 and 1985 distributed across 70 independent watersheds (see Henriques-Silva *et al.*, 2012, for complete details). In this study, each watershed was considered as a metacommunity. The number of sites (local communities) and species varied quite a lot across watersheds (between 21 and

280 lakes and 17 and 50 species). In order to contrast the two spatial models (MSSN and MEM), we simply contrasted their R^2_{adj} across all 70 watersheds.

RESULTS

Simulated data

Figure 2 summarizes the simulation results comparing our MSSN and MEM approaches for sample sizes of 100 communities (Appendix S2 contains the results based on 150 communities). First, on the basis of explained variance (average R^2_{adj} values across all simulated communities), it is clear that our novel approach outperforms MEM (Fig. 2). Second, both

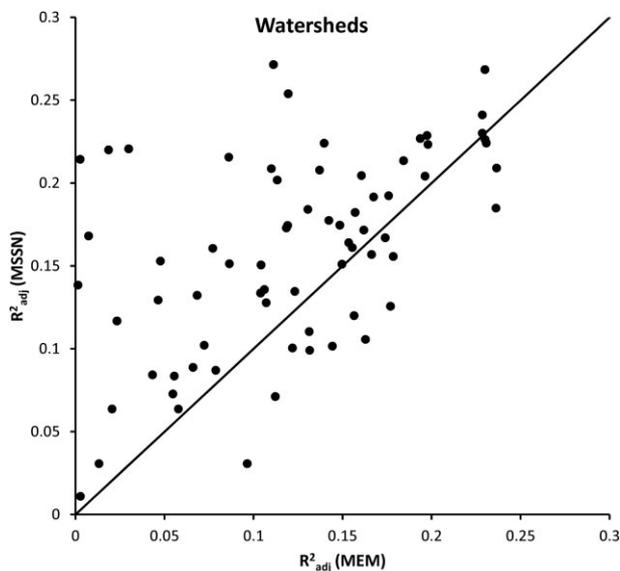


Figure 3 Contrast between our multi-species spatial network (MSSN) method and the Moran's eigenvector maps (MEM) method on the basis of adjusted R^2 values obtained from real ecological data sets (70 fish metacommunities).

methods are, as expected, sensitive to the level of spatial autocorrelation in which low spatial ranges reduce the ability of both methods to detect spatial patterns in species community composition (Fig. 2). Third, both methods are sensitive to the sample size, in that large samples increase the performance of both methods (compare Fig. 2 with Appendix S2). Fourth, both methods were sensitive to the level of random (non-spatial) replacement of communities; greater levels of non-spatial noise (contrast 20% with 0% replacement; Fig. 2) decreased the performance of both methods. Fifth, not only did our framework generate models with superior predictive power to MEM in explaining spatial variation in metacommunities (Fig. 2), it also showed greater statistical power. Finally, the Type I error of our framework leads to a valid test and is in fact below the expected value (Fig. 2, range = 0).

Real ecological data

Figure 3 contrasts the R^2_{adj} values for the MSSN and MEM frameworks across 70 watersheds. The results clearly show the advantage of our spatial network method over MEM in detecting spatial patterns in a large set of real data. For the majority of watersheds (56 watersheds out of 70, or 80% of the watersheds), the metacommunity models based on the MSSN approach had greater explanatory power (R^2_{adj}) than those based on MEM. The fact that our spatial network had inferior predictive power for 16 watersheds may mean that these watersheds do not have strong spatial signatures.

An additional advantage of our framework is that it can make inferences about the strength of the connectivity levels across local communities. Figure 4 contrasts two metacommunities (watersheds) in terms of the spatial patterns of local community connec-

tivity. Each circle in Fig. 4 represents a lake within the watershed and the size of the circles shows the amount of connectivity with other lakes. Moreover, the size of the circles are proportional to the number of links, both inner (immigration) events and outer (emigration) events. On average, lakes within watershed 2 are better connected than lakes within watershed 1 (Fig. 4).

DISCUSSION

Inferring complex spatial patterns in real metacommunities is a daunting task given the logistic limitations in following individuals across a wide range of taxa and geographic locations. Moreover, the processes shaping metacommunities may have been historical and much beyond the temporal scope of the empirical data on species distributions (Leibold *et al.*, 2010; Layeghifard *et al.*, 2012). The framework we introduced here is intended to detect spatial variability in metacommunities and represent them as spatial networks.

Given the challenges of directly measuring spatial patterns within metacommunities, the proposed framework offers several advantages. First, it provides a parallel framework to those used in metapopulation models (Hanski, 1994; Dunham & Rieman, 1999; Hartel *et al.*, 2010). However, our measure of connectivity is based on functions that represent spatial distributions of occupied versus non-occupied sites for multiple species (homologous to metapopulation metrics; see Bender *et al.*, 2003, and Prugh, 2009, for reviews). Second, our method infers connectivity matrices that can be then used as spatial predictors in multiple species modelling frameworks. This is akin to single-species modelling that uses metapopulation connectivity metrics to estimate site isolation (e.g. Dunham & Rieman, 1999; Prugh, 2009). It also allows us to estimate how well connected (hot spots) or disconnected (cold spots) local communities are. The levels of connectivity can also themselves be contrasted against predictors such as patch size (e.g. lake area) and environmental predictors.

The final advantage, that our MSSM framework is capable of detecting patterns of connectivity that are not necessarily spatialized, deserves some additional attention because it relates to the way that connectivity metrics for metapopulations and ours (metacommunity) make inferences about dispersal. Although we commonly assume that the signatures of dispersal are spatialized, communities that are spatially close may harbour different species and, conversely, communities that are spatially distant may assemble similar species. In these cases, one would have a reduced ability to infer dispersal solely on the basis of the spatial structure of species distributions. That is the reason why metapopulation metrics use incidence information across the landscape and our method uses community similarity across the metacommunity. However, as in metapopulation metrics, our MSSN framework also weights community similarity in relation to geographic distance by considering a compromise between spatial arrangement and information on community similarity. In other words, if there is a high level of similarity between two communities that are quite far apart they will not be considered as connected as if the same two communities were near to

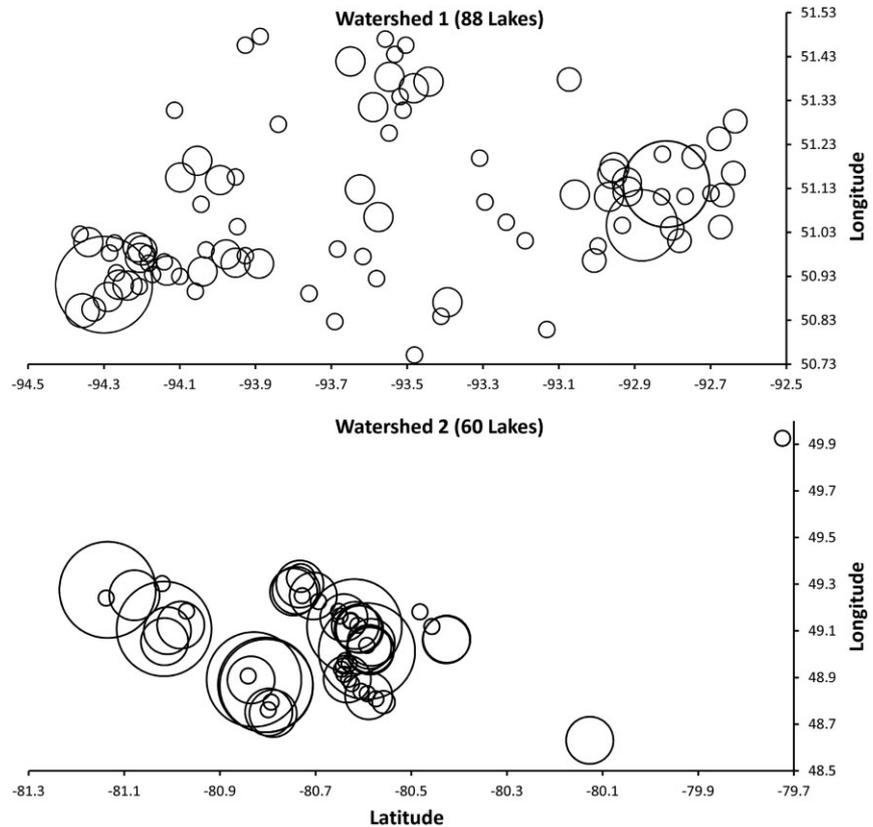


Figure 4 Bubble plot maps for lakes of two fish metacommunities (watersheds) representing their levels of connectivity to other lakes within each watershed. Lakes are plotted according to their geographic positioning. The size of the circles represents the levels of connectivity for any particular lake.

each other. This is an important issue when studying metapopulations and metacommunities inhabiting environmentally heterogeneous landscapes, especially those composed of species that are good dispersers but which have strong environmental preferences. In this case, species can get anywhere (mass effect perspective; Leibold *et al.*, 2004) but are sorted according to the type of environment (species-sorting perspective). It follows that metapopulation metrics and our metacommunity framework may infer strong dispersal dynamics across occupied sites (especially in the case where optimal patches are also spatially structured), whereas in fact the major factor is strong environmental filtering instead. One way to separate these two hypotheses is to use a variation partitioning approach (Borcard *et al.*, 1992; Peres-Neto *et al.*, 2006) in which environment and our MSSN predictors are contrasted against each other. In cases where species have strong environmental affinities (optimal environments are highly spatially structured) and are not limited by dispersal, environmental predictors and our MSSN predictors should covary strongly and serve as an indication that our dispersal predictors are confounded by environment. Note that this is not an issue particular to our method but one of natural landscapes not being able to provide 'perfect' orthogonal designs to estimate the influence of environmental filtering versus dispersal (i.e. variation in species habitat optima is also spatially structured).

Our simulation and real data applications clearly show the features and advantages of our framework over a method that is widely used to depict spatial patterns in metacommunities (i.e. MEM). The main advantage of our method is that it integrates both geographic

and species composition information to infer and explain spatial heterogeneity in species distributions. By integrating both sources of information we can also characterize non-spatialized dispersal patterns. Although the geographic distance between patches is a fundamental component of any landscape, species dispersal behaviour also plays a key role in shaping the spatial structure of metacommunities. In cases where species distributions closely follow the geographic arrangement of the habitat patches in the landscape, species dispersal can be ignored in practice. However, in reality, species dispersal patterns are much more complex and our network framework aims to infer such patterns directly.

Our approach is very general and flexible. One needs to set up a spatial network and then add extra branches on the basis of similarities in species composition between local communities. Any algorithm can be used to build the spatial network and any metric can be used to measure similarity in species composition; as such, one can certainly consider additional ways to measure the interactions between vertices (local communities). For instance, Brooks *et al.* (2008) described discrete (e.g. different threshold distances in which sites are considered connected or unconnected) and continuous edge weights (the strength of interaction between pairs of vertices) that could certainly be used when building matrix H . Because different H matrices can be built on the basis of these different choices and each regressed against the same response matrix (here species distribution matrix), we can contrast the R^2_{adj} of these different spatial matrices and retain the model that best explains multi-species distributions. In this paper, our goal was to describe a general framework in which we reconcile a geographic spatial network with a species composition network to infer patterns of spatial connectivity

within networks. As such, we made a decision to use a simple set of approaches instead of comparing a large number of potential extensions of our method. Note, however, that many potential extensions of our framework were provided in the text and in Appendix S1. That said, our implementation (geographic distance and Jaccard similarity) was clearly superior to current approaches for measuring spatial patterns in metacommunities. The main application of the spatial approach developed here was intended to be in modelling multi-species distributions, which is an extremely common goal in community and metacommunity ecology. However, unlike other landscape graph approaches (e.g. Urban & Keitt, 2001; Brooks *et al.*, 2008), our spatial connectivity framework can only be estimated when species distribution information is readily available. Note, however, that it is important to ensure that different methodological procedures have correct Type I error so that performance of different methodological implementations is indeed due to power. Therefore, different methodological applications should undergo a statistical evaluation applying similar procedures to those used in this study.

We certainly hope that ecologists find our approach innovative, useful and intuitive. Our MSSN framework is quite flexible and can easily incorporate different types of functions to infer spatial proximity (e.g. nonlinear functions), different types of indices to infer community similarity and different types of algorithms to infer cluster of sites and dispersal directionality. All the applications mentioned above could be performed by changing the input data, using our framework in conjunction with other methods or applying slight modifications to the framework. As such, we expect that our MSSN method will become a valuable addition to the spatial ecologists' toolbox and find many interesting applications in metacommunity studies, landscape ecology and conservation biology.

ACKNOWLEDGEMENTS

This study was supported by the FQRNT (Fonds de Recherche sur la Nature et les Technologies du Québec) team research grant to V. Makarenkov and P. Peres-Neto, and an FQRNT PhD grant to M. Layeghifard.

REFERENCES

Banašek-Richter, C., Bersier, L.-F., Cattin, M.-F., Baltensperger, R., Gabriel, J., Merz, Y., Ulanowicz, R.E., Tavares, A.F., Williams, D.D., Ruitter, P.C., Winemiller, K.O. & Naisbit, R.E. (2009) Complexity in quantitative food webs. *Ecology*, **90**, 1470–1477.

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.

Bender, D.J., Tischendorf, L. & Fahrig, L. (2003) Using patch isolation metrics to predict animal movement in binary landscapes. *Landscape Ecology*, **18**, 17–39.

Bodin, O. & Norberg, J. (2007) A network approach for analyzing spatially structured populations in a fragmented landscape. *Landscape Ecology*, **22**, 31–44.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

Brooks, C.P. (2006) Quantifying population substructure: extending the graph-theoretic approach. *Ecology*, **87**, 864–872.

Brooks, C.P., Antonovics, J. & Keitt, T.H. (2008) Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *The American Naturalist*, **172**, 149–159.

Bunn, A.G., Urban, D.L. & Keitt, T.H. (2000) Landscape connectivity: a conservation application of graph theory. *Journal of Environmental Management*, **59**, 265–278.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Dray, S. (2011) A new perspective about Moran's coefficient: spatial autocorrelation as a linear regression problem. *Geographical Analysis*, **43**, 127–141.

Dray, S., Legendre, P. & Peres-Neto, P. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Dray, S., Péliissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P.R., Bellier, E., Bivand, R., Blanchet, F.G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J. & Wagner, H. (2012) Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, **82**, 257–275.

Dunham, J.B. & Rieman, B.E. (1999) Metapopulation structure of bull trout: influences of physical, biotic, and geometrical landscape characteristics. *Ecological Applications*, **9**, 642–655.

Felsenstein, J. (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, MA.

Foltête, J.-C., Clauzel, C., Vuidel, G. & Tournant, P. (2012) Integrating graph-based connectivity metrics into species distribution models. *Landscape Ecology*, **27**, 557–569.

Griffith, D.A. & Peres-Neto, P.R. (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, 2603–2613.

Gucht, K.V., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., Jeppesen, E., Conde-Porcuna, J.-M., Schwenk, K., Zwart, G., Degans, H., Vyverman, W. & Meester, L.D. (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 USA*, **104**, 20404–20409.

Hanski, I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, **63**, 151–162.

Hanski, I. & Ovaskainen, O. (2000) The metapopulation capacity of a fragmented landscape. *Nature*, **404**, 755–758.

Hartel, T., Nemes, S., Öllerer, K., Cogalniceanu, D., Moga, C. & Arntzen, J.W. (2010) Using connectivity metrics and niche modeling to explore the occurrence of the northern crested newt (Amphibia, Caudata) in a traditionally managed landscape. *Environmental Conservation*, **37**, 195–200.

- Henriques-Silva, R., Lindo, Z. & Peres-Neto, P.R. (2012) A community of metacommunities: exploring patterns in species distributions across large geographical areas. *Ecology*, **94**, 627–639.
- Holyoak, M., Leibold, M.A. & Holt, R.D. (2005) *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press, Chicago.
- Huston, M.A. (1999) Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. *Oikos*, **86**, 393–401.
- Jacobson, B. & Peres-Neto, P.R. (2010) Quantifying and disentangling dispersal in metacommunities: how close have we come? How far is there to go? *Landscape Ecology*, **25**, 495–507.
- Jordán, F., Magura, T., Tóthmérész, B., Vasas, V. & Ködöböcz, V. (2007) Carabids (Coleoptera: Carabidae) in a forest patchwork: a connectivity analysis of the Bereg Plain landscape graph. *Landscape Ecology*, **22**, 1527–1539.
- Krause, A.E., Frank, K.A., Mason, D.M., Ulanowicz, R.E. & Taylor, W.W. (2003) Compartments revealed in food-web structure. *Nature*, **426**, 282–285.
- Layeghifard, M., Peres-Neto, P.R. & Makarenkov, V. (2012) Using directed phylogenetic networks to retrace species dispersal history. *Molecular phylogenetics and evolution*, **64**, 190–197.
- Legendre, P. & Legendre, L. (2012) *Numerical ecology*. Elsevier, Amsterdam.
- Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.
- Leibold, M.A., Economo, E.P. & Peres-Neto, P. (2010) Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology Letters*, **13**, 1290–1299.
- Luczkovich, J.J., Borgatti, S.P., Johnson, J.C. & Everett, M.G. (2003) Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology*, **220**, 303–321.
- Meyer, K.M., Ward, D., Wiegand, K. & Moustakas, A. (2008) Multi-proxy evidence for competition between savanna woody species. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 63–72.
- Mihaescu, R., Levy, D. & Pachter, L. (2009) Why neighbor-joining works. *Algorithmica*, **54**, 1–24.
- Peres-Neto, P. & Cumming, G. (2010) A multi-scale framework for the analysis of fish metacommunities. *Community ecology of stream fishes: concepts, approaches, and techniques* (ed. by G.B. Gido and D.A. Jackson), pp. 235–262. American Fisheries Society, New York.
- 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.
- Prugh, L.R. (2009) An evaluation of patch connectivity measures. *Ecological Applications*, **19**, 1300–1310.
- Rubio, L. & Saura, S. (2012) Assessing the importance of individual habitat patches as irreplaceable connecting elements: an analysis of simulated and real landscape data. *Ecological Complexity*, **11**, 28–37.
- Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.
- Urban, D. & Keitt, T. (2001) Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**, 1205–1218.
- Zeller, K.A., McGarigal, K. & Whiteley, A.R. (2012) Estimating landscape resistance to movement: a review. *Landscape Ecology*, **27**, 777–797.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Example of calculation for the multi-species spatial network (MSSN) approach.

Appendix S2 Contrast in performance between the MSSN versus MEM (Moran's eigenvector maps) on the basis of metacommunities based on 150 simulated communities.

BIOSKETCHES

Mehdi Layeghifard (PhD) is currently a post-doctoral fellow at the Department of Cell and Systems Biology, University of Toronto. He is interested in evolutionary biology, biogeography and bioinformatics, with emphasis on improving and developing new analytical computational methods.

Pedro R. Peres-Neto (PhD) is a community quantitative ecologist with an interest in examining the roles of multiple ecological factors such as species-level traits (e.g. morphology, dispersal capacity, life history, phenotypic integration), habitat choice, landscape structure and species interactions in driving species distributions and community structure.

Vladimir Makarenkov (PhD) works in the fields of computational biology, software engineering and mathematical classification. His research interests include design and development of bioinformatics applications and databases, investigation of the evolution of molecular sequences, reconstruction of phylogenetic trees and reticulation networks and study of algorithmic time complexity.

Editor: Sean Connolly