The role of environmental and spatial processes in structuring native and non-native fish communities across thousands of lakes

Sapna Sharma, Pierre Legendre, Miquel De Cáceres and Daniel Boisclair

S. Sharma (sapna.sharma23@gmail.com), P. Legendre, M. De Cáceres and D. Boisclair, Dépt des Sciences Biologiques, Univ. de Montréal, C.P. 6128 succursale Centre-ville, Montréal, QC H3C 3J7, Canada. (Present address of SS: Center for Limnology, 680 N. Park Street, Univ. of Wisconsin-Madison, Madison, WI 53706, USA.)

Quantifying the role of spatial patterns is an important goal in ecology to further understand patterns of community composition. We quantified the relative role of environmental conditions and regional spatial patterns that could be produced by environmental filtering and dispersal limitation on fish community composition for thousands of lakes. A database was assembled on fish community composition, lake morphology, water quality, climatic conditions, and hydrological connectivity for 9885 lakes in Ontario, Canada. We utilized a variation partitioning approach in conjunction with Moran's Eigenvector Maps (MEM) and Asymmetric Eigenvector Maps (AEM) to model spatial patterns that could be produced by human-mediated and natural modes of dispersal. Across 9885 lakes and 100 fish species, environmental factors and spatial structure explained approximately 19% of the variation in fish community composition. Examining the proportional role of spatial structure and environmental conditions revealed that as much as 90% of the explained variation in native species assemblage composition is governed by environmental conditions. Conversely on average, 67% of the explained variation in non-native assemblage composition can be related to human-mediated dispersal. This study highlights the importance of including spatial structure and environmental conditions when explaining patterns of community composition to better discriminate between the ecological processes that underlie biogeographical patterns of communities composed of native and non-native fish species.

Enhancing our knowledge of the spatial scales at which ecological patterns are generated has been an important goal in ecology (Levin 1992) and remains a vital challenge for community ecologists (Borcard et al. 2004). Communities are structured by the interaction of species with their abiotic and biotic environments at different spatial scales (Jackson et al. 2001, Borcard et al. 2004). Traditionally, only environmental conditions have been used to explain patterns of community composition. However, recently the metacommunity framework (see Leibold et al. 2004 for a review) has highlighted the importance of considering spatial patterns when describing community composition (Cottenie 2005, Beisner et al. 2006, Leprieur et al. 2009).

Spatial structure in community composition is the result of many ecological processes acting simultaneously. Two kinds of spatial structures can be distinguished: induced spatial dependence and spatial autocorrelation. Induced spatial dependence is defined as the spatial structure of the response variables which exists due to the spatial structure of the explanatory variables. The traditional environmental control model (Whittaker 1956, Hutchinson 1957) suggests that induced spatial dependence is generated by the spatial structure of environmental factors acting on biological communities through species-habitat associations (Legendre

et al. 2009). Spatial autocorrelation is the second type of spatial structure and is generated by the species themselves from biotic processes such as growth, dispersal, and mortality (Legendre and Legendre 1998, Fortin and Dale 2005). Dispersal limitation is a particularly strong factor that can lead to spatial autocorrelation by producing aggregated patterns in species composition (Hubbell 2001, Legendre et al. 2009).

Species dispersal has important implications for biodiversity and community composition (Mouquet and Loreau 2002). As rates of dispersal increase, the similarity among sites is also expected to increase resulting in decreased beta diversity (Shurin et al. 2009). For fish species, natural dispersal of species can arise as a result of postglacial isostatic movements (Legendre and Legendre 1984, Mandrak and Crossman 1992) and movement of fishes through hydrological connectivity (Olden et al. 2001). Human-mediated dispersal of fishes can occur by authorized, prohibited, and accidental introductions by anglers and government agencies (Crossman 1991). Currently species movement is occurring at an unprecedented rate (Ricciardi 2007) and the dispersal rate for non-native species is much higher than native species owing to multiple introductions in geographically distant areas (Wilson et al. 2009). As climate change is expected to

exacerbate the rate of spread of non-native species (Dukes and Mooney 1999, Sharma et al. 2007), there is currently a need to quantify patterns of community composition before further increased rates of dispersal.

Prior to the advent of variation partitioning (Borcard et al. 1992) and novel spatial statistical methods, such as Moran's Eigenvector Maps (MEM; Borcard and Legendre 2002, Borcard et al. 2004, Dray et al. 2006) and Asymmetric Eigenvector Maps (AEM; Blanchet et al. 2008), the quantification of the relative role of environmental conditions, and non-directional and directional spatial components acting on patterns of community composition was not possible. In this study, we provide a quantification of the importance of environmental conditions and spatial patterns (Cottenie 2005, Shurin et al. 2009) using MEM, AEM, and variation partitioning methods in structuring fish communities for ten thousand lakes across a landscape. The distance-based MEM eigenfunctions computed here were formerly called Principal Coordinates of Neighbour Matrices (PCNM; Borcard and Legendre 2002, Dray et al. 2006). More specifically, the primary objective was to quantify the relative role of environmental conditions and spatial patterns that could arise from niche-based processes including environmental filtering (Keddy 1992) and dispersal limitation (Hubbell 2001), such as natural and human-mediated modes of dispersal. The secondary objective was to assess the differences in patterns of community composition between native and non-native fish species across several drainage basins, thereby determining if patterns of community composition are consistent across the landscape and for assemblage types. The analyses of different assemblage types elucidate an important unanswered question as to whether native and non-native species assemblages are structured similarly (Sax et al. 2007). We hypothesize that native non-sport fish communities would be primarily structured by environmental conditions, whereas fish communities composed of non-native fish species would be predominately structured by human-mediated dispersal.

Methods

Data acquisition

We gathered data on fish communities, lake morphology, water quality, and climatic conditions for 9885 lakes in Ontario. Fish community and lake information were obtained from the Ontario Habitat Inventory database collected by the Ontario Ministry of Natural Resources between 1957 and 1987 and aimed to obtain a representative sample of a gradient of lakes in each region across Ontario (Table 1). Fish community composition was primarily assessed using gillnet surveys and comprised 100 fish species and over 67 000 records of fish occurrences (Dodge et al. 1985, Minns 1986). A suite of environmental variables (Dodge et al. 1985) was collected in conjunction with the fish community surveys during a single sampling event in the summer. These environmental variables consisted of lake geography (latitude, longitude, elevation), morphology (i.e. surface area, maximum depth, mean depth, perimeter), water quality (i.e. total dissolved solids, conductivity, Secchi depth,

Table 1. Minimum, maximum, and mean values of environmental conditions observed for the 9885 lakes in Ontario collected between 1957 and 1987.

Variable	Minimum	Maximum	Mean
Area (ha)	0.1	79771.6	311.5
Maximum depth (m)	0.2	213.5	17
Mean depth (m)	0.1	50.3	5.6
Elevation (m)	42	875	341.3
Secchi depth (m)	0.2	36.6	3.7
Total dissolved solids (mg L^{-1})	2	828	59.1
Mean annual air temperature ($^{\circ}$C)	−1.9	8.7	2.7
Mean annual precipitation (mm)	1.5	3.3	2.3

surface water temperatures, pH, and oxygen concentrations), and climate (i.e. growing degree days) (Dodge et al. 1985).

We obtained climate data, specifically mean monthly air temperatures and mean annual precipitation, from the Intergovernmental Panel on Climate Change (IPCC) Data Distribution Centre as 1961–1990 averages. Climate data were interpolated from meteorological stations using thin-plate splines and summarized on a $0.5^{\circ} \times 0.5^{\circ}$ latitude/longitude grid. In addition, hydrological network data were acquired from Natural Resources Canada in the form of the National Scale Frameworks Hydrology database (<http://geogratis.cgdi.gc.ca>). The database represented information on surface water, river flow direction, and connectivity compiled at a scale of 1:1 000 000.

Data manipulation

Our data required several transformations to satisfy the assumptions of statistical analyses. Environmental variables with low geographic coverage across the spatial extent, that is environmental variables which were not sampled in many lakes, were removed prior to analyses. Furthermore, environmental variables exhibiting high levels of multicollinearity were removed prior to analyses as high multicollinearity between environmental variables may reduce the ability to correctly quantify model parameters (Legendre and Legendre 1998). The remaining lake variables (surface area, maximum depth, mean depth, elevation, Secchi depth, and total dissolved solids) were log-transformed to increase linearity of relationships and avoid extreme values. Species data were subjected to the Hellinger transformation prior to multivariate analyses. This transformation consists of expressing each species occurrence as a proportion of the sum of all species occurrences in each lake, and taking the square root of the resulting value. The square-root step of the transformation decreases the importance of the most abundant species. This transformation is recommended for use in linear ordinations (Legendre and Gallagher 2001). Furthermore, very rare species (defined as the presence of a species in one lake in the drainage basin) were removed prior to analyses. Rare species can have a disproportionate effect on multivariate analyses as they contribute little or nothing to understanding general community relationships (Jackson and Harvey 1989).

We divided our Ontario database into five drainage basins: 1) Lakes Ontario, Erie, and St. Clair; 2) Lake Huron; 3) Lake Superior; 4) St. Lawrence and Ottawa Rivers; and 5) Nelson River (Fig. 1; Table 2). The dataset
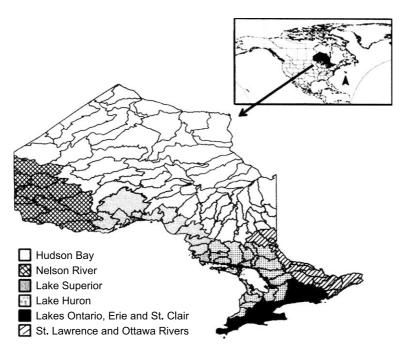
Figure 1. Map of the drainage basins and tertiary watersheds in Ontario. Lines represent tertiary watershed boundaries.

was divided to ascertain general patterns of community composition across drainage basins and to circumvent limitations in computational ability due to large sample sizes since the routines of eigenvalue decomposition we used could not handle matrices larger than 5500×5500. We omitted lakes from the Hudson Bay drainage basin from the analyses because of the increased likelihood of an under-estimation of species diversity due to small sampling effort in this region (Minns and Moore 1995). In general, lakes south of 52°N latitude have been well sampled while those north of 52°N latitude have been sparsely sampled (Mandrak and Crossman 1992).

Data analyses were performed at two hierarchical levels: community and assemblage types. Community analyses were performed on all species found in the drainage basin, with the exception of rare species. We defined four assemblage types: 1) native non-sport fishes, 2) native sport fishes, 3) native bait fishes, and 4) non-native fishes (Table 3). Only established fishes were classified using biological and economic information provided by Scott and Crossman (1973) and Mandrak and Crossman (1992). We used the categorization of non-native species from Mandrak and Crossman (1992) who defined non-native fishes to Ontario as 'a species intentionally or unintentionally introduced into non-native ecosystems' (Mandrak and Crossman 1992). Species that were rare or known to have low detectability and

sampling effort were not categorized into fish assemblage types (Scott and Crossman 1973).

Data analyses

Modelling symmetrical spatial structure

Principal Coordinates of Neighbour Matrices (PCNM) was developed by Borcard and Legendre (2002) to quantify symmetric spatial structure, that is, spatial patterns that do not incorporate an assumption of directionality. PCNMs are a special form of MEM eigenfunctions, as mentioned above. These MEM variables, or eigenfunctions, are obtained from a spectral decomposition of a truncated distance matrix of the spatial relationships among sampling locations. In our case, we used direct overland linear distances between the lakes (i.e. the centre of each lake) as input spatial relationships. The specifics of the analyses are as follows. First the response variable is detrended (if necessary) to remove the linear trend in the response data. A linear trend is an indication of spatial structure acting at a broader scale than the sampling extent. It may also obscure other structure in the dataset that could be more optimally modelled by MEM variables. The residuals of the linear regression are retained for subsequent analyses. Secondly, distance-based MEM eigenfunctions are constructed. This is done by computing the Euclidean distance among the sites. In the matrix of Euclidean distances, one finds the smallest distance that maintains connections between all sites along a minimum spanning tree; this value is called the truncation distance. The distances between lakes that are larger than the truncation distance are replaced by 4 times the truncation distance. The modified, truncated distance matrix is referred to as the neighbour matrix. A principal coordinates analysis (PCoA) is conducted on the neighbour matrix. Two-thirds of the eigenvalues produced by the PCoA will be positive. These eigenvalues will be referred to as MEM variables and represent the spatial structure in the dataset. The MEM variables will be in the form of a series

Table 2. Summary of the number of lakes and fish species in the five Ontario drainage basins analysed in this study.

Drainage basin	Number of lakes	Number of fish species
Lakes Ontario, Erie and St. Clair	605	60
Lake Huron	2242	71
Lake Superior	1047	46
St. Lawrence and Ottawa River	1655	62
Nelson River	1676	66

Table 3. Classification of fishes into four assemblage types: native non-sport fishes, native sport fishes, native bait fishes, and non-native fishes.

Non-sport fishes	Sport fishes
Lake sturgeon *Acipenser fulvescens*	Channel catfish *Ictalurus punctatus*
Longnose gar *Lepisosteus osseus*	Northern pike *Esox lucius*
Gizzard shad *Dorosoma cepedianum*	Muskellunge *Esox masquinongy*
Longnose sucker *Catostomus catostomus*	Lake herring *Coregonus artedi*
Yellow bullhead *Ameiurus natalis*	Lake whitefish *Coregonus clupeaformis*
Brown bullhead *Ameiurus nebulosus*	Brook trout *Salvelinus fontinalis fontinalis*
Round whitefish *Prosopium cylindraceum*	Lake trout *Salvelinus namaycush*
Burbot *Lota lota*	Smallmouth bass *Micropterus dolomieu*
Banded killifish *Fundulus diaphanus*	Largemouth bass *Micropterus salmoides*
Rock bass *Ambloplites rupestris*	Black crappie *Pomoxis nigromaculatus*
Pumpkinseed *Lepomis gibbosus*	Yellow perch *Perca flavescens*
Bluegill *Lepomis macrochirus*	Walleye *Sander vitreus*
Sauger *Sander canadensis*	
Northern redbelly dace *Phoxinus eos*	
White sucker *Catostomus commersonii*	
Common bait fishes	**Non-native fishes**
Common shiner *Luxilus cornutus*	Alewife *Alosa pseudoharengus*
Pearl dace *Margariscus margarita*	Goldfish *Carassius auratus*
Golden shiner *Notemigonus crysoleucas*	Common carp *Cyprinus carpio*
Emerald shiner *Notropis atherinoides*	Rainbow smelt *Osmerus mordax*
Blacknose shiner *Notropis heterolepis*	Rainbow trout *Oncorhynchus mykiss*
Spottail shiner *Notropis hudsonius*	Brown trout *Salmo trutta*
Blacknose dace *Rhinichthys atratulus*	Splake *Salvelinus fontinalis* x *Salvelinus namaycush*
Creek chub *Semotilus atromaculatus*	
Brook silverside *Labidesthes sicculus*	

of sine waves with decreasing periods that are orthogonal to one another. Finally, a linear analysis (such as redundancy analysis if the response data are multivariate and multiple linear regression if response data are univariate) is performed on the response data with a set of predictor variables (such as MEM and environmental variables, Fig. 2a; Borcard and Legendre 2002, Borcard et al. 2004). MEM eigenfunctions describe symmetric spatial structures at all spatial scales that can be expressed by the sampling design. The first MEM variables model broad spatial structures, and subsequent MEM variables represent smaller spatial patterns. The last eigenfunctions accommodate fine-scale spatial structures (see Borcard and Legendre 2002, Borcard et al. 2004 for details).

Modelling asymmetrical spatial structure

Asymmetric Eigenvector Maps (AEM) model directional spatial structures (Blanchet et al. 2008). The method first requires an identification of the asymmetrical links between sites. Spatial eigenfunctions are then obtained from the singular value decomposition of the (column-centred) matrix of asymmetric links. Similar to MEM, asymmetrical spatial eigenfunctions can be subsequently used as spatial descriptors in linear regression methods (see Blanchet et al. 2008 for details). More specifically, first the response variable is detrended (if necessary) to remove the linear trend in the response data. The residuals of the linear regression are retained for subsequent analyses. Secondly, AEM eigenfunctions are constructed. This is completed by first constructing a connectivity matrix which summarises the connections between lakes and hydrological networks. Directionality is incorporated into the connection matrix by specifying the direction of the connection between lakes and hydrological networks. Subsequently a principal components analysis (PCA) or singular value decomposition (SVD) is conducted on the connection diagram to create a series of spatial eigenfunctions. Finally, a linear analysis is performed on the response data with a set of predictor variables (such as AEM and environmental variables, Fig. 2b; Blanchet et al. 2008).

Partitioning environmental and spatial variation

We identified the relative importance of environmental conditions (lake morphology, water quality, and climate), symmetric spatial structure (MEM variables), and asymmetric spatial structure (AEM variables) by partitioning the variation explained by each component using redundancy analyses (Borcard et al. 1992). The shared variation between environmental variables and spatial descriptors are produced by induced spatial dependence generated by the spatial structure of environmental factors acting indirectly on biological communities. We summarized the variation explained by each component by calculating the adjusted R^2. While R^2 is biased and increases with the number of explanatory variables even if they are random, R^2_{adj} provides unbiased estimates of the variation of the response data explained by the explanatory variables (Peres-Neto et al. 2006).

The fraction of the variation explained by regression of MEM variables alone represents symmetric spatial structures of the fish community that could arise from non-directional spatial processes. We hypothesize that human-mediated dispersal may be explaining symmetrical spatial patterns, as the anthropogenic releases of fish are not following hydrological networks (Litvak and Mandrak 1993), and are therefore unlikely to generate asymmetrical spatial patterns following a hydrological network. The use of MEM variables serves two purposes. First, the unique fraction explained by MEM variables are used as a surrogate of human-mediated dispersal and suggest how fishes are being moved across the landscape. Second, the shared fraction explained by MEM and environmental variables represent spatially structured environmental conditions and suggest how environmental conditions are structured across the landscape and responding to the fish community.

AEM could have been used to describe the river network interconnections among lakes within each drainage basin. However, due to the paucity of detailed data available on hydrological networks for the 9885 lakes in the Ontario Habitat Inventory database, we only identified the hydrological connections between the tertiary watersheds for each
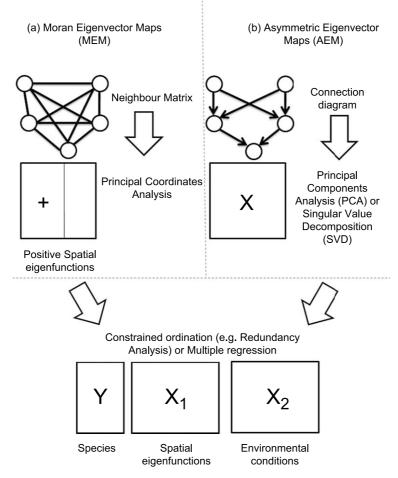
Figure 2. Description of distance-based Moran's Eigenvector Maps (MEM, formally called Principal Coordinates of Neighbour Matrices, PCNM) and Asymmetric Eigenvector Maps (AEM) analyses.

drainage basin. Thus we used the AEM variables to model asymmetric watercourse connections at a broader spatial scale (i.e. tertiary watersheds) than the MEM or environmental variables, which were representing spatial patterns acting at a finer spatial scale (i.e. for each lake). We hypothesize that broad-scale natural dispersal of fishes between tertiary watersheds following hydrological networks through glacial history (Mandrak and Crossman 1992) could produce asymmetrical spatial patterns. If present, such patterns led to an increase in the fraction fitted by AEM variables alone. However, to demonstrate the utility of AEM analyses to quantify the role of hydrological connections among lakes, we present an example on a subset of Ontario lakes with detailed knowledge on river network interconnections among lakes (Supplementary material Appendix 1).

We summarised the association between fish assemblage types and the drainage basin to which the assemblage type belonged to determine if the environmental and spatial patterns exhibited by fish assemblages were similar because of drainage basin location or broad characteristics of assemblage types. To this end, we performed a principal components analysis (PCA) on the adjusted percent variation explained solely by MEM, AEM, and environmental variables. All data manipulation and statistical analyses were performed in the R-language environment (R Development Core Team 2009).

Results

Across 9885 lakes and 100 fish species, environmental conditions, symmetric spatial structure, and asymmetric spatial structure together explained on average ca 19% of the variation (62% unadjusted variation) in fish community composition (Supplementary material Appendix 2). The average adjusted explained variation differed for the assemblage types: native non-sport fishes (15.6%), sport fishes (27.4%), bait fishes (14.2%), and non-native fishes (10.4%; Supplementary material Appendix 2). The amount of variation explained ranged from −0.5% (for non-native fish community in the St. Lawrence/Ottawa River drainage basin) to 36.8% (for sport fish community in Lake Superior) (Supplementary material Appendix 2). The model describing gizzard shad *Dorosoma cepedianum* occurrence in the lower Great Lakes explained the most variation of any model, explaining over 76% of the variation. Over 70% of the explained variation was attributed to environmental factors (both solely and spatially structured). Furthermore, 10% of the variation

in gizzard shad occurrence was uniquely explained by asymmetrical dispersal between watersheds suggesting that the gizzard shad is naturally a highly mobile species (Supplementary material Appendix 3–7).

The following results summarise the variation explained as a percentage of the total fitted variation. Environmental conditions including induced spatial dependence accounted for ca 78% of the explained variation in the fish community. MEM and AEM variables (including those that quantified induced spatial dependence) accounted for 59 and 37% of the explained variation respectively in fish community composition (Fig. 3a).

The native non-sport fish assemblage was predominately structured by environmental variables (including induced spatial dependence), which accounted for ca 92% of the explained variation on average across drainage basins. 5% of the variation in the fish community was uniquely explained by MEM variables representing symmetrical spatial patterns (Fig. 3b). The native sport fish assemblage was primarily structured by environmental variables, accounting for ca 82% of the explained variation, across drainage basins. MEM variables uniquely accounted for 14% of the explained variation in the sport fish community (Fig. 3c). The native bait fish assemblage is also structured by environmental variables accounting for ca 68% of the explained variation in the assemblage across drainage basins. Furthermore, 25% of the explained variation in the native bait fish assemblage was uniquely accounted for by MEM variables (Fig. 3d). Non-native species composition was primarily explained by the unique contribution of symmetric spatial variables which accounted for 67% of the explained variation across drainage basins (Fig. 3e; Supplementary material Appendix 2).
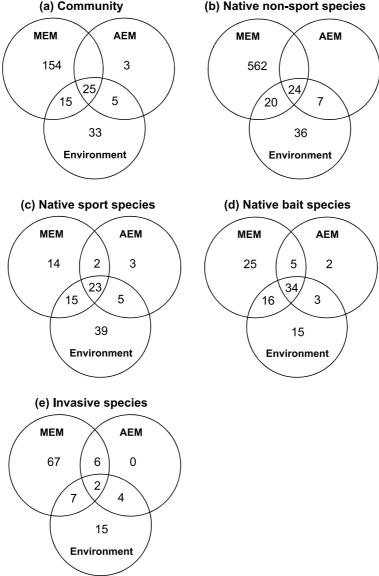

Figure 3. Fractions of adjusted percent variation (R^2_{adj}) explained for fish community composition (a) and each fish assemblage type, native non-sport fish species (b), native sport fish species (c), native bait fish species (d), and non-native fish species (e) on average for all drainage basins by the set of predictor variables (MEM = variation explained by pure symmetric spatial structure; AEM = variation explained by pure asymmetric spatial structure; ENV = variation explained by pure environmental conditions) based on variation partitioning analyses.

Spatially structured environmental variables such as lake morphology, water quality, and climate explained only 30% of the variation, and the AEM variables tended to explain negligible amounts of variation in the non-native fish community.

On average across drainage basins the PCA biplot confirmed that assemblage types exhibited similar patterns across drainage basins (Fig. 4). Non-native fish assemblages appear to be highly related to symmetrical spatial processes modelled by the MEM variables and largely unrelated to environmental variables. In general, AEM variables do not play a large role in structuring fish assemblages at the broad spatial scale. The sport fish assemblage appears highly related to environmental variables, such as lake morphology, water quality, and climate. The native non-sport fish assemblage had low percentages of variation explained by MEM variables. Finally, the bait fish community was located south-east of the origin on the ordination biplot, in the sector opposite to the sole influence of environmental variables suggesting that the bait fish assemblage was primarily related to spatially structured environmental conditions (Fig. 4).

Discussion

To our knowledge, this is the first study providing a framework to quantify the relative importance of symmetrical spatial patterns using overland distances, asymmetrical
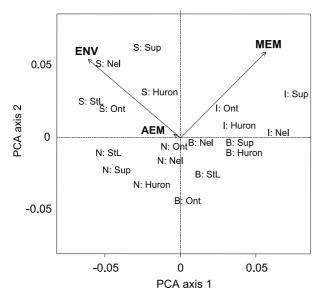

Figure 4. Principal components analysis (PCA) summarizing the variation explained (adjusted R^2) for each assemblage type (I = non-native fishes; S = sport fishes; N = native non-sport fishes; and B = bait fishes) and drainage basin (Ont = Lakes Ontario, Erie, and St. Clair; Hur = Lake Huron; Sup = Lake Superior; StL = St. Lawrence and Ottawa Rivers; Nel = Nelson River). Points located close together on the PCA axis are associated with each other, whereas points found at opposite ends of the PCA axis are unlikely to be associated with each other. The length of the arrow reflects the importance of the variable. The first and second PCA axes summarized ca 58.5 and 40.3% of the variation respectively in variation explained by each assemblage type and drainage basin.

spatial patterns using broad-scale watercourse distances, and environmental conditions in structuring fish communities composed of native and non-native species across thousands of lakes. The ecological patterns that were quantified could be produced by environmental filtering (Keddy 1992) and dispersal limitation (Hubbell 2001) processes. In this study, we used environmental variables that summarized lake morphology, water quality, and climatic conditions acting on each lake to assess the role of environmental filters explaining patterns of community composition. Asymmetrical spatial patterns modelled by AEM variables were used to assess the broad-scale natural watercourse dispersal of species through hydrological connectivity and finally, direct overland linear distances between lakes were used to quantify symmetric spatial patterns using MEM variables (Beisner et al. 2006). We hypothesized that human-mediated dispersal may be explaining symmetrical spatial patterns, as anthropogenic releases of fish are not following hydrological networks (Litvak and Mandrak 1993).

This study contributes to the current debate in the biogeography and ecological literature about whether the distributions of native and non-native species follow similar ecological rules. Some studies have suggested that native species exhibit similar biogeographical patterns as non-native species (Sax 2001, LaSorte and McKinney 2006) while others have found that the biogeography of native communities is as a result of environmentally filtering processes and the biogeography of non-native communities is a result of human-mediated processes (Olden et al. 2008, Leprieur et al. 2009). Contrasting the fractions explained by environmental conditions and symmetric spatial structure alone for native fishes and non-native fishes indicates that native and non-native fish assemblages across broad landscapes are not driven by similar ecological rules. This finding confirms the result from freshwater studies which suggest that the biogeography of native fishes is predominantly, but not exclusively driven by environmental filtering processes and that the biogeography of non-native fishes is primarily governed by human-mediated dispersal limitation (Jeschke and Strayer 2006, Olden et al. 2008, Leprieur et al. 2009). As such, the broad-scale patterns of freshwater fish distribution result from the combined effects of environmental filtering processes and historical contingency factors, such as dispersal limitation due to geographical barriers, differential rates of extinction and speciation, and differential rates of post-glacial colonization (Jackson and Harvey 1989, Olden et al. 2001, Leprieur et al. 2009), in addition to biotic interactions such as the presence of predators (Englund et al. 2009).

Environmental variables and spatial patterns accounted for ca 20% of the variation (62% unadjusted variation) in community composition for one hundred fish species across ten thousand lakes. The inclusion of spatial variables explained on average an additional 21% of variation in fish community composition across drainage basins. The fraction explained by spatial structure increases for fishes known to disperse through human-mediated means, such as non-native species. Furthermore, the inclusion of spatial patterns explains 46% of the variation that is shared between environmental and spatial variables in fish community composition.

Role of environmental conditions

At a community-wide level, environmental variables accounted for ca 80% of the explained variation in fish community composition, although 60% of this variation was spatially-structured. The shared variation between environmental variables and spatial descriptors is produced by induced spatial dependence suggesting that environmental conditions play a large role in patterns of community composition. It further suggests that inclusion of variables quantifying spatial structure is imperative to quantifying the factors structuring communities (Cottenie 2005, Shurin et al. 2009).

Specifically, we found that there were strong differences in the importance of environmental conditions on community assemblage composition, such that 87% of the variation for native non-sport vs 30% of the variation for non-native assemblages was explained by environmental conditions. All native fish species, whether categorized as non-sport, sport or bait fishes, were predominately structured by environmental conditions. This suggests that niche-determining and environmental filtering processes, such as suitable local abiotic and biotic conditions affect community composition, and thereby are largely influencing the biogeography of native fish assemblages (Leprieur et al. 2009). Conversely, environmental filtering processes were not strongly influencing the biogeography of non-native assemblages as non-native fishes tend to be tolerant of a wide range of environmental conditions (Ehrlich 1989, Lodge 1993). Additionally, the interaction between anthropogenic activity and environmental conditions may contribute to the low influence of environmental variables on the composition of non-native fish communities.

Role of non-directional spatial patterns (MEM variables)

The MEM approach allowed us to model spatial variables that are structured non-directionally. The shared variation between MEM and environmental variables is likely acting at broader spatial scales accounting for environmental heterogeneity and induced spatial dependence, whereas dispersal and biotic interactions such as competition and predation are likely acting at finer spatial scales (Borcard et al. 2004). The fraction of the fish community solely explained by MEM variables may be attributed to spatial processes occurring over the landscape, such as those resulting from dispersal of fishes via anthropogenic releases either via intentional or accidental introduction of fishes, stocking, and bait bucket transfers (Crossman 1991). We found that MEMs accounted for a significant fraction of variation, suggesting that symmetrical spatial processes are important in structuring fish communities. This was confirmed in a study of river networks in the Mississippi-Missouri basin, Muneepeerakul et al. (2008) found that the dispersal of most fish is symmetrically structured and not biased asymmetrically.

There were strong differences in the importance of symmetrical spatial patterns on assemblage composition (5% for native non-sport vs 67% for non-native) suggesting that non-native species are likely to be dispersed through

human-mediated means over geographical barriers (Rahel 2007, Olden et al. 2008, Leprieur et al. 2009). Furthermore, as the amount of anthropogenic interference increases with respect to human-mediated movement of fishes among lakes, the fraction quantifying symmetrical spatial patterns increases. Only 5% of the variation in native non-sport fish assemblage composition was explained by pure symmetrical spatial structure. The fraction increased to 14% for the native sport fish assemblage and may be reflecting dispersal of sport fishes by anthropogenic means, such as stocking. For example, intentional stocking programs have altered the distribution and abundance of sport fishes in Ontario, such as smallmouth bass, largemouth bass *Micropterus salmoides*, brook trout *Salvelinus fontinalis timagamiensis* (Minns and Moore 1995), and lake trout (Johnson et al. 1977). The stocking of smallmouth bass into Ontario lakes to provide angling opportunities (Johnson et al. 1977) led to unexpected consequences. The introduction of smallmouth bass can lead to the homogenization of fish fauna as cyprinid species are lost (MacRae and Jackson 2001) and to reduced growth, survival, and fecundity of lake trout populations (Vander Zanden et al. 1999). The fraction quantifying pure symmetrical spatial pattern increased to 25% for the native bait fish community. In Ontario, there is a large amount of dispersal of bait fishes occurring by anthropogenic means, such as bait bucket transfers when anglers release their live bait, such as blacknose dace, creek chub *Semotilus atromaculatus*, and golden shiner *Notemigonus crysoleucas* following fishing (Minns 1989) and this is reflected in the increased variation explained by the MEM variables.

Finally, 67% of the variation in the non-native fish assemblage was explained by symmetrical spatial structure. The large fraction of variation explained by MEM variables may be reflecting the unprecedented rate of human-mediated dispersal (Ricciardi 2007) with species introduced in multiple locations and by multiple modes of introduction (Wilson et al. 2009). Authorized and prohibited intentional introductions of non-native fishes continue to occur in Ontario for nutritional, recreational, nostalgic, aquarium, biological control, forage, and bait fish purposes (Crossman 1991) and this results in higher dispersal rates for non-native species compared to native species, and the establishment of non-native species in regions well beyond their native range (Wilson et al. 2009).

The current rate of human-mediated dispersal is unprecedented (Ricciardi 2007) with species introduced in multiple locations and by multiple modes of introduction (Wilson et al. 2009). This results in higher dispersal rates for invasive species compared to native species, and the establishment of invasive species in regions well beyond their native range (Wilson et al. 2009). Following initial introduction, successful invasive species tend to possess natural mechanisms of rapid dispersal permitting subsequent successful invasions of other waterbodies (Vanderploeg et al. 2002).

Role of non-directional spatial patterns (AEM variables)

The AEM approach modelled directional spatial structure through watercourse connections through hydrological

networks, and may be accounting for processes such as broad-scale natural dispersal of fishes between tertiary watersheds through post-glaciation corridors and hydrological connectivity (permanent headwater connexions, temporary connexions due to freshets). It should be reiterated that AEM variables were constructed at a broader spatial scale (i.e. tertiary watersheds) than the MEM or environmental variables, which were representing patterns occurring at finer spatial scales (i.e. for each lake). Furthermore, information on barriers to dispersal was not included due to the nature of the databases. As such, the AEM variables did not generally explain much variation in community composition. This may also be attributed to the fact that most fish tend to disperse locally, although a few species may disperse long distances (Jackson et al. 2001, Muneepeerakul et al. 2008). However, Legendre and Legendre (1984) found that headwater connexions, modelled by AEM in the present study, were a major factor explaining present-day freshwater fish distributions in Québec.

Detailed analyses by species indicated that the occurrence of some fish species, such as gizzard shad and blacknose dace *Rhinichhthys atratulus*, were significantly explained by AEM variables, suggesting that AEM variables can be powerful at modelling highly mobile species that disperse naturally through watersheds. If detailed hydrological connectivity information was available for all 9885 lakes, AEM variables would most likely have explained more variation in structuring the fish community via the finer-scale natural dispersal of fishes. In a subset of Ontario lakes where detailed hydrological connectivity was available, AEM variables explained ca 75% of the variation in fish community composition suggesting that natural dispersal plays a large role in structuring fish communities (Supplementary material Appendix 1).

There was no significant role of asymmetrical spatial processes in explaining non-native community composition, indicating that non-native species distribution is not dictated by broad-scale dispersal through hydrological networks. This may be because species introductions are a relatively recent phenomenon so there may not have been enough time for non-native species to disperse across tertiary watersheds. However, highly naturally mobile species may have the ability to naturally disperse long distances through hydrological networks that may increase the fraction explained by asymmetrical spatial processes in the future.

Conclusions

This study reflects the importance of including both environmental conditions and spatial components when identifying patterns of community composition to better understand the ecological processes that underlie distributional patterns of community composition. As such, we provide a variation partitioning framework using novel spatial statistical approaches that enables the direct quantification of the roles of environmental conditions and spatial components on community composition. Our results support recent work indicating that environmental filtering and spatial processes, such as dispersal, are important in explaining patterns of community composition (Cottenie

2005, Beisner et al. 2006, Leprieur et al. 2009). Furthermore, we quantify the difference in the patterns explaining community composition for native and non-native species, which had been a previously demonstrated need (Sax et al. 2007). Across a broad landscape, our study suggests that biogeographical patterns of native and non-native species do not follow similar ecological rules. Rather the distribution of native species is primarily determined by environmental filtering processes, whereas the distribution of non-native species is largely a consequence of dispersal limitation likely due to anthropogenic interference (Leprieur et al. 2008, 2009, Olden et al. 2008). As human-mediated dispersal is the primary determinant of non-native species distribution, it is crucial to alter and implement policy, management, and conservation strategies to limit the potential dispersal of non-native species.

References

Beisner, B. E. et al. 2006. The role of environmental and spatial processes in structuring lake communities from bacteria to fish. – Ecology 87: 2985–2991.

Blanchet, F. G. et al. 2008. Modelling directional spatial processes in ecological data. – Ecol. Model. 215: 325–336.

Borcard, D. and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. – Ecol. Model. 153: 51–68.

Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.

Borcard, D. et al. 2004. Dissecting the spatial structure of ecological data at multiple scales. – Ecology 85: 1826–1832.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. – Ecol. Lett. 8: 1175–1182.

Crossman, E. J. 1991. Introduced freshwater fishes: a review of the North American perspective with emphasis on Canada. – Can. J. Fish. Aquat. Sci. 48: 46–57.

Dodge, D. P. et al. 1985. Manual of instructions: aquatic habitat inventory surveys. – Ontario Ministry of Natural Resources, Fisheries Policy Branch, Toronto, ON, Canada.

Dray, S. et al. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). – Ecol. Model. 196: 483–493.

Dukes, J. S. and Mooney, H. A. 1999. Does global change increase the success of biological invaders? – Trends Ecol. Evol. 14: 135–139.

Ehrlich, P. R. 1989. Attributes of invaders and the invading processes: vertebrates. – In: Drake, J. A. et al. (eds), Biological invasions: a global perspective. Wiley, pp. 315–328.

Englund, G. et al. 2009. Predation leads to assembly rules in fragmented fish communities. – Ecol. Lett. 12: 663–671.

Fortin, M.-J. and Dalel, M. R. T. 2005. Spatial analysis: a guide for ecologists. – Cambridge Univ. Press.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. – Princeton Univ. Press.

Hutchinson, G. E. 1957. Concluding remarks. – Cold Spring Harbor Symp. Quant. Biol. 22: 415–427.

Jackson, D. A. and Harvey, H. H. 1989. Biogeographic associations in fish assemblages: local versus regional processes. – Ecology 70: 1472–1484.

Jackson, D. A. et al. 2001. What controls who is where in freshwater fish communities – the roles of biotic, abiotic, and spatial factors. – Can. J. Fish. Aquat. Sci. 58: 157–170.

Jeschke, J. M. and Strayer, D. L. 2006. Determinants of vertebrate invasion success in Europe and North America. – Global Change Biol. 12: 1608–1619.

Johnson, M. G. et al. 1977. Limnological characteristics of Ontario lakes in relation to associations of walleye (*Stizostedion vitreum vitreum*), northern pike (*Esox lucius*), lake trout (*Salvelinus namaycush*), and smallmouth bass (*Micropterus dolomieui*). – J. Fish. Res. Board Can. 34: 1592–1601.

Keddy, P. A. 1992. Assembly and response rules: two goals for predictive community ecology. – J. Veg. Sci. 3: 157–164.

La Sorte, F. A. and McKinney, M. L. 2006. Compositional similarity and the distribution of geographical range size for assemblages of native and non-native species in urban floras. – Divers. Distrib. 12: 679–686.

Legendre, P. and Legendre, V. 1984. Postglacial dispersal of freshwater fishes in the Quebec peninsula. – Can. J. Fish. Aquat. Sci. 41: 1781–1802.

Legendre, P. and Legendre, L. 1998. Numerical ecology. – Elsevier.

Legendre, P. and Gallagher, E. 2001. Ecologically meaningful transformations for ordination of species data. – Oecologia 129: 271–280.

Legendre, P. et al. 2009. Partitioning beta diversity in a subtropical broad-leaved forest of China. – Ecology 90: 663–674.

Leibold, M. A. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. – Ecol. Lett. 7: 601–613.

Leprieur, F. et al. 2008. Fish invasions in the world's river systems: when natural processes are blurred by human activities. – PloS Biol. 6: e28.

Leprieur, F. et al. 2009. Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. – J. Biogeogr. 36: 1899–1912.

Levin, S. A. 1992. The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.

Litvak, M. K. and Mandrak, N. E. 1993. Ecology of freshwater baitfish use in Canada and the United States. – Fisheries 18: 6–13.

Lodge, D. M. 1993. Biological invasions: lessons for ecology. – Trends Ecol. Evol. 8: 133–137.

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

Mandrak, N. E. and Crossman, E. J. 1992. Postglacial dispersal of freshwater fishes in Ontario. – Can. J. Zool. 70: 2247–2259.

Minns, C. K. 1986. A model of bias in lake selection for survey. – Can. Tech. Rep. Fish. Aquat. Sci. No. 1496.

Minns, C. K. 1989. Factors affecting fish species richness observed in Ontario's inland lakes. – Trans. Am. Fish. Soc. 118: 533–545.

Minns, C. K. and Moore, J. E. 1995. Factors limiting the distributions of Ontario's freshwater fishes: the role of climate and other variables, and the potential impacts of climate change. – In: Beamish, R. J. (ed.), Climate change and northern fish populations. Canadian Special Publication of Fisheries and Aquatic Sciences 121, pp. 137–160.

Mouquet, N. and Loreau, M. 2002. Community patterns in source-sink metacommunities. – Am. Nat. 162: 544–557.

Muneepeerakul, R. et al. 2008. Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. – Nature 453: 220–223.

Olden, J. D. et al. 2001. Spatial isolation and fish communities in drainage lakes. – Oecologia 127: 572–585.

Olden, J. D. et al. 2008. Species invasions and the changing biogeography of Australian freshwater fishes. – Global Ecol. Biogeogr. 17: 25–37.

Peres-Neto, P. R. et al. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. – Ecology 87: 2614–2625.

R Development Core Team 2009. R: a language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna, Austria, <www.R-project.org>.

Rahel, F. 2007. Biogeographic barriers, connectivity, and biotic homogenization: it's a small world after all. – Freshwater Biol. 82: 696–710.

Ricciardi, A. 2007. Are modern biological invasions an unprecedented form of global change? – Conserv. Biol. 21: 329–336.

Sax, D. F. 2001. Latitudinal gradients and geographic ranges of exotic species: implications for biogeography. – J. Biogeogr. 28: 139–150.

Sax, D. F. et al. 2007. Ecological and evolutionary insights from species invasions. – Trends Ecol. Evol. 22: 464–471.

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

Sharma, S. et al. 2007. Will northern fish populations be in hot water because of climate change? – Global Change Biol. 13: 1052–1064.

Shurin, J. B. et al. 2009. Spatial autocorrelation and dispersal limitation in freshwater organisms. – Oecologia 159: 151–159.

Vander Zanden, M. J. et al. 1999. Stable isotope evidence for the food web consequences of species invasions in lakes. – Nature 401: 464–467.

Vanderploeg, H. A. et al. 2002. Dispersal and emerging ecological impacts of Ponto-Caspian species in the Laurentian Great Lakes. – Can. J. Fish. Aquat. Sci. 59: 1209–1228.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. – Ecol. Monogr. 26: 1–80.

Wilson, J. R. U. et al. 2009. Something in the way you move: dispersal pathways affect invasion success. – Trends Ecol. Evol. 24: 136–144.

Supplementary material (Appendix E6811 at <www.oikos. ekol.lu.se/appendix>). Appendix 1–7.