Received: 20 July 2017 | Revised: 11 May 2018 | Accepted: 29 May 2018
DOI: 10.1111/geb.12781

RESEARCH PAPER

WILEY Global Ecology and Biogeography A Journal of Macroecology

Similarity in spatial structure constrains ecosystem relationships: Building a macroscale understanding of lakes

Jean-Francois Lapierre[1,2] | **Sarah M. Collins**[3] | **David A. Seekell**[4,5] | **Kendra Spence Cheruvelil**[6,7] | **Pang-Ning Tan**[8] | **Nicholas K. Skaff**[6] | **Zofia E. Taranu**[9] | **C. Emi Fergus**[10] | **Patricia A. Soranno**[6]

[1]Département de sciences biologiques, Université de Montréal, Pavillon Marie-Victorin, Montréal, Québec, Canada
[2]Groupe de Recherche Interuniversitaire en Limnologie et en Environnement, Aquatique (GRIL), Montréal, Québec, Canada
[3]Center for Limnology, University of Wisconsin, Madison, Wisconsin
[4]Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden
[5]Climate Impacts Research Centre, Umeå University, Abisko, Sweden
[6]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan
[7]Lyman Briggs College, Michigan State University, East Lansing, Michigan
[8]Department of Computer Science & Engineering, Michigan State University, East Lansing, Michigan
[9]Department of Biology, University of Ottawa, Ottawa, Ontario, Canada
[10]National Research Council, US Environmental Protection Agency, Corvallis, Oregon

Correspondence
Jean-Francois Lapierre, Département de sciences biologiques, Université de Montréal, Pavillon Marie-Victorin, CP 6128, succursale Centre-ville, Montréal, QC H3C 3J7, Canada.
Email: jean-francois.lapierre.1@umontreal.ca

FUNDING INFORMATION
NSF, Grant/Award Number: EF-1065786, EF-1065818 and DBI-1401954; Fonds de Recherche du Québec en Nature et Technologies; National Science and Engineering Research Council of Canada; USDA National Institute of Food and Agriculture, Hatch Project, Grant/Award Number: 176,820; Knut and Alice Wallenberg Foundation

Abstract

Aim: We aimed to measure the dominant spatial patterns in ecosystem properties (such as nutrients and measures of primary production) and the multi-scaled geographical driver variables of these properties and to quantify how the spatial structure of pattern in all of these variables influences the strength of relationships among them.

Location and time period: We studied > 8,500 lakes in a 1.8 million km^2 area of Northeast U.S.A. Data comprised 10-year medians (2002–2011) for measured ecosystem properties, long-term climate averages and recent land use/land cover variables.

Major taxa studied: We focused on ecosystem properties at the base of aquatic food webs, including concentrations of nutrients and algal pigments that are proxies of primary productivity.

Methods: We quantified spatial structure in ecosystem properties and their geographical driver variables using distance-based Moran eigenvector maps (dbMEMs). We then compared the similarity in spatial structure for all pairs of variables with the correlation between variables to illustrate how spatial structure constrains relationships among ecosystem properties.

Results: The strength of spatial structure decreased in order for climate, land cover/use, lake ecosystem properties and lake and landscape morphometry. Having a comparable spatial structure is a necessary condition to observe a strong relationship between a pair of variables, but not a sufficient one; variables with very different spatial structure are never strongly correlated. Lake ecosystem properties tended to have an intermediary spatial structure compared with that of their main drivers, probably because climate and landscape variables with known ecological links induce spatial patterns.

Main conclusions: Our empirical results describe inherent spatial constraints that dictate the expected relationships between ecosystem properties and their geographical drivers at macroscales. Our results also suggest that understanding the spatial scales at which ecological processes operate is necessary to predict the effects of multi-scaled environmental changes on ecosystem properties.

KEYWORDS
climate, ecosystem, lake, landscape, macroscales, Moran eigenvector maps, spatial autocorrelation, spatial scale, spatial structure

1 | INTRODUCTION

Ecosystems are highly heterogeneous over space and influenced by processes that operate at local to continental scales. There is a growing need to understand how regional to global changes will affect ecosystem processes and the services they provide at the macroscale (Clark et al., 2001; Heffernan et al., 2014; Qiu & Turner, 2013). In response to this need, frameworks and approaches have been developed to study ecosystem processes and patterns across spatial extents (Heffernan et al., 2014; Levy et al., 2014; Rose et al., 2017; Soranno et al., 2014). Such multi-scaled analyses, however, have rarely been grounded in empirical studies, owing in part to a lack of fine-resolution data on ecosystem properties in the many thousands of ecosystems across regions and continents. Given that sampling often occurs above or below the scale of interest (Anselin, 2001; Fortin & Dale, 2005), it remains challenging to understand how both fine- and broad-scaled ecological and geographical drivers may induce spatial structure in ecosystem properties.

Recent efforts to compile large, multi-scaled datasets combined with better methods for quantifying spatial patterns are facilitating ecological studies of these questions at macroscales (e.g., Sharma et al., 2015; Soranno et al., 2017), but these opportunities present analytical and conceptual challenges. One of these challenges is to understand and quantify the role of spatial structure in the relationships between ecosystem properties and their geographical (landscape and climatic) drivers. Spatial structure, or the shape and strength of spatial patterns, is often quantified using various measures of spatial autocorrelation (Fortin & Dale, 2005). Spatial autocorrelation has long been recognized as a problem that may inflate error in statistical models (Hoeting, 2009; Legendre, 1993) but is increasingly being used as a meaningful tool to estimate how different types of variables may induce spatial patterns in ecosystem properties. For example, spatial autocorrelation patterns in stream chemistry suggested that reactive nutrients were locally structured in response to in-stream uptake, whereas conservative tracers of terrestrial inputs were more broadly structured at the whole watershed level (McGuire et al., 2014). In lakes, studies to date have shown evidence for regional structure (i.e., spatial autocorrelation among lakes within regions) in some key lake chemistry and biotic variables. Between-lake variance plateaued for the concentrations of: chlorophyll *a* and nutrients at *c*. 100 km in Michigan, U.S.A. (Cheruvelil, Soranno, Bremigan, Wagner, & Martin, 2008), dissolved organic carbon at *c*. 300 km in Sweden (Seekell et al., 2014), and chlorophyll *a* and bacterial respiration at *c*. 200–300 km in boreal Quebec (Lapierre, Seekell, & del Giorgio, 2015). In contrast, climate variables in the boreal region were spatially autocorrelated for sites distant up to 10,00 km (Lapierre et al., 2015). Thus, there is evidence that different categories of ecosystem properties may have contrasting spatial structure, but the spatial structure of a comprehensive suite of climate, landscape and ecosystem properties has never been quantified simultaneously within an individual study at macroscales. Therefore, although spatial structure has been relatively well characterized for distributions and interactions of organisms (Gotelli, Graves, & Rahbek, 2010; McGill, 2010; Rahbek, 2005; Wiens, 1989), the role of spatial structure in understanding the spatial patterns in ecosystem matter and energy at broad spatial extents is largely unknown.

Lakes are ideally suited to the study of patterns across different spatial scales because their boundaries are well defined and because it is possible to measure how they receive and process material originating from the lake itself, the overlying atmosphere and the surrounding landscape. Furthermore, lakes are ecological and biogeochemical hotspots in the landscape (McClain et al., 2003; Strayer & Dudgeon, 2010), and it is crucial to understand the flow of matter and energy at local, regional and continental scales through these important ecosystems by developing a stronger understanding of the spatial patterns of the factors that control these flows. Recent studies, taken together, suggest that lakes respond to diverse and interacting geographical drivers operating at multiple spatial scales, such as continental to global climate, regional land cover and land use and local catchment properties (Lapierre, et al., 2015; O'Reilly et al., 2015; Seekell et al., 2014). The ecological or mathematical constraints underlying such broad-scale ecosystem patterns, however, have only been studied implicitly and have not been elucidated.

Our goal in the present study was to gain a better understanding of how spatial structure varies across ecosystem properties and their geographical drivers, and how that spatial structure influences relationships among ecosystem and geographical variables. We used an empirical approach that takes advantage of emerging large integrated databases of ecological systems at broad scales. We use LAGOS-NE (LAke multi-scaled GeOSpatial and temporal), a large, multi-scaled and multi-themed database of> 8,000 lakes in a subcontinental spatial extent (Soranno et al., 2017, 2015). Our specific objectives were as follows: (a) to quantify the spatial structure in ecosystem properties and their geographical drivers; (b) to examine whether the strength of the relationship between ecosystem and geographical variables is related to similarity in spatial structure; and (c) to compare these empirical findings with analyses of simulated data to assess the generality of our results. Identifying the spatial principles that link ecosystem patterns from fine scales to macroscales should improve the ability of ecologists to understand and forecast the effects of environmental changes on ecosystem functioning across regions and continents.

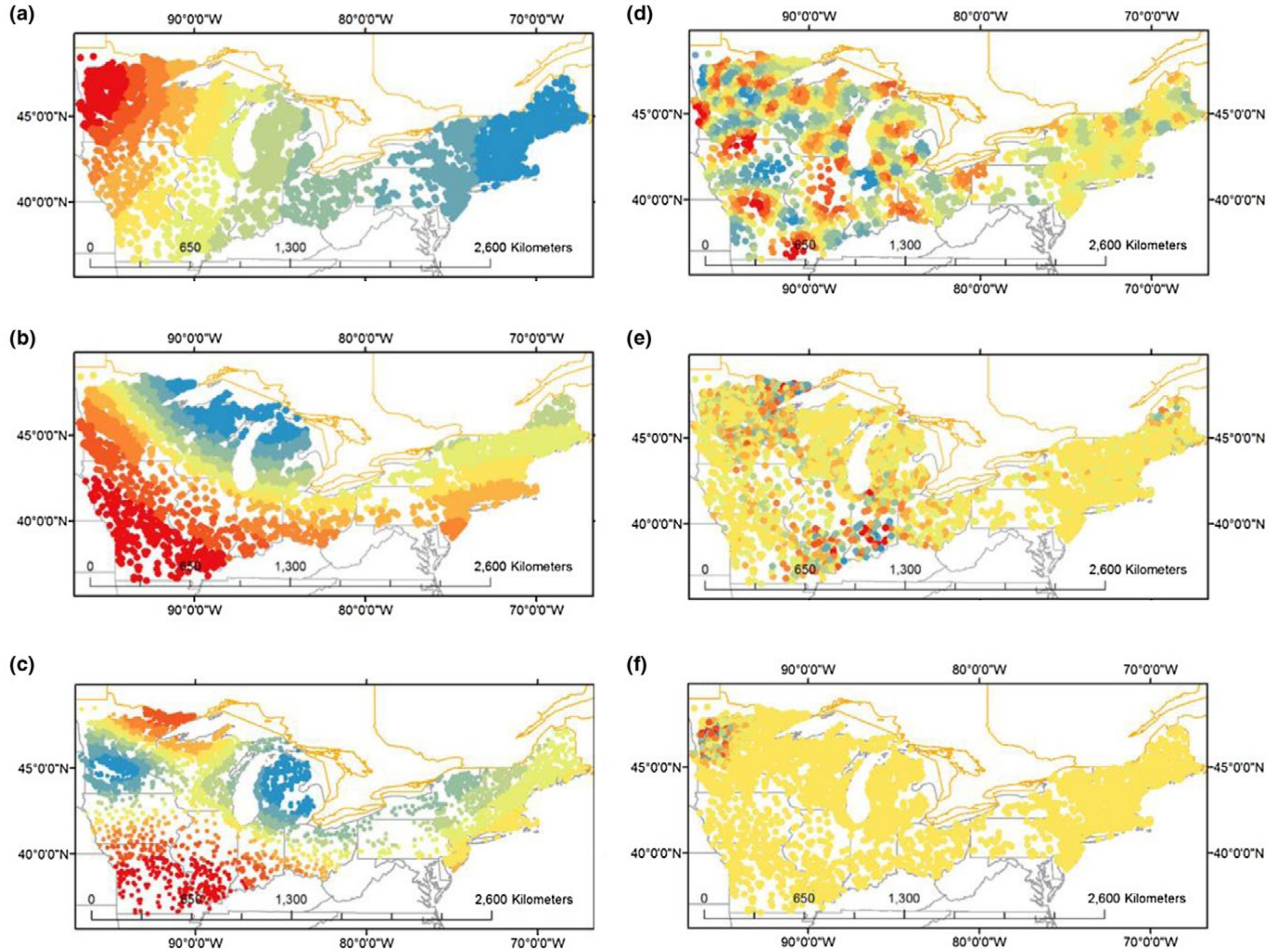

FIGURE 1 Examples of distance-based Moran eigenvector maps (dbMEMs) used to explain spatial patterns in study variables. The dbMEMs are shown in decreasing order in terms of spatial structure. dbMEM numbers 1, 3, 6, 100, 1,000 and 3,000 (out of 8,743 total) are shown. For each panel, the colours represent the amplitude of the sine wave of the dbMEM, the "S-value", at each site. Thus, the dbMEMs roughly correspond to wavelengths of 3,200, 1,400, 800, 200, 10 and 0 km, respectively. Sites with a similar S-value colour are the most similar to each other and red versus blue values are inversely correlated [Colour figure can be viewed at wileyonlinelibrary.com]

2 | METHODS

2.1 | Study extent and data sources

We analysed climate, landscape and lake ecosystem properties in a subcontinental spatial extent (*c*. 1.8 million km^2) in the temperate Midwest and Northeast regions of the U.S.A. (Figure 1). This spatial extent includes wide environmental and climatic gradients that vary several-fold, including: hydrology (e.g., mean runoff 44–762 mm/year), land use/land cover (e.g., agricultural land use 0%–89%), climate (e.g., mean annual precipitation and temperature 566–1,376 mm/year and 2.4–14.5°C, respectively), lake morphometry (e.g., lake area 0.01–1,237 km^2) and nutrient concentrations (e.g., total phosphorus 2–998 µg/L).

We used the LAGOS-NE database, including data modules for geographical data (LAGOS$_{GEO}$, version 1.03; LAGOS$_{LOCUS}$, version 1.01) and lake ecosystem data (LAGOS$_{LIMNO}$, version 1.054.1) (Soranno et al., 2017). We used a subset of the LAGOS-NE database that includes any lake that has at least one record for the ecosystem variables we considered (n = 8,744). We quantified the spatial structure in ecosystem and geographical variables from two data sources. The first data source was from geographical data that were summarized by hydrological units (HU), which are hierarchically nested stream watersheds based on United States Geological Survey (USGS) 1:24,000 scale topographic maps. In this analysis, we summarized the ecological property data that were not directly associated with a lake or its watershed at the smallest available HU extents, i.e. at the HU12 scale (median area = 78 km^2), in which each lake was assigned to a HU12 polygon. We also characterized the lake watersheds (a crucial ecological property of lakes) by measuring catchment slope, watershed area and watershed land use/cover. We associated the limnological data for the 8,744 lakes with the corresponding geographical data in their respective HU12 so that each lake had ecological properties calculated at the scale of HU12 s, the lake itself and the watershed of the lake. Note that some HU12 s contain more than one lake so that they would be assigned identical HU12 data, but not the same lake watershed data.

TABLE 1 Relationships of distance-based Moran eigenvector maps to contrasting spatial structure and climate, landscape and lake ecosystem properties

	MEM number																
Variable	Latitude	Longitude	1	2	3	4	5	6	7	8	9	10	11	13	14	15	18	Cumulative r^2
Mean annual precipitation	0.07		0.82					0.02										0.91
Mean annual temperature	0.91		0.02															0.93
Runoff		0.88			0.02			0.02										0.92
SO_4 deposition	0.68		0.1			0.02		0.04	0.04									0.88
Mixed forest	0.11	0.21			0.29			0.13									0.02	0.76
N deposition	0.03	0.28			0.18			0.27	0.04					0.02				0.82
Evergreen	0.31	0.22						0.1		0.02	0.02							0.67
Agriculture			0.05	0.14	0.31			0.1		0.02								0.62
Baseflow					0.35			0.21		0.04								0.60
Forested wetlands					0.41			0.07										0.48
Urban	0.22						0.04	0.11	0.02	0.06		0.04						0.49
Wetlands	0.33				0.05					0.03						0.02		0.43
Pasture				0.04	0.26			0.04							0.02			0.36
TP	0.12	0.18				0.04		0.02		0.05								0.41
Chl a					0.19	0.04	0.05	0.03										0.31
Catchment slope		0.04	0.27					0.02		0.04								0.37
Secchi depth	0.05	0.05		0.02	0.18		0.03											0.33
TN		0.11		0.09			0.03	0.04		0.03			0.02					0.32
Lake depth (maximum)							0.07	0.02										0.09
Lake area	0.05						0.03											0.08
Lake perimeter							0.03											0.03
Watershed area	0.01																	0.01
Approximate MEM wavelength (km)			3,200	1,600	1,400	1,100	1,000	800	700	600	550	500						

Notes. Variables are shown with decreasing spatial structure from top to bottom. The intensity of the red colour is proportional to the strength of the r^2 between a particular distance-based Moran eigenvector map (dbMEM) and a variable. Green-labelled variables represent climate and landscape properties, light-blue labelled variables represent common lake ecosystem properties for which there was sufficient spatial coverage, and deep-blue variables represent lake and catchment morphometry. Only the top eight dbMEMs with the most predictive power are included in the table. A few other dbMEMs explained little amounts of variation in only one ecosystem variable but are not included here for clarity. The full table is shown as Supporting Information Appendix S3.

Geographical data were available for all lakes, but sample size varied for the lake ecosystem data [n = 3,452, 1,902, 8,004 and 4,733 for total phosphorus (TP), total nitrogen (TN), Secchi depth (Secchi) and chlorophyll a (Chl a), respectively]. For lake ecosystem data, we used the median of summer values (15 June 15–15 September) over a period from 2002 to 2011, from which most of the data came. Specific information about the geographical data is provided by Soranno et al. (2017). The database used in this analysis is available (Lapierre, et al., 2017).

2.2 | Quantifying spatial structure using distance-based Moran eigenvector maps

We quantified spatial structure as the cumulative r^2 between each ecosystem property variable and the distance-based Moran eigenvector maps (dbMEMs) as described below. First, we built dbMEMs (Dray, Legendre, & Peres-Neto, 2006) based on geographical coordinates of the 8,744 lakes for which we had limnological data, using the adespatial package (Dray et al., 2016) in R (R Core Team, 2017). Distance-based Moran eigenvector maps are the orthogonal eigenvectors of a spatial weight matrix (i.e., geographical distance matrix) whose corresponding eigenvalues represent decreasing levels of spatial autocorrelation (the Moran's I coefficients, which define how a variable correlates with itself between two different locations and are commonly used to study spatial patterns in ecology; Legendre & Legendre, 2012). Positive spatial correlation in the data translates into positive values of I; negative correlation produces negative values. There are n − 1 potential dbMEMs (eigenvectors), half of which model the positive spatial correlation (positive Moran's I) and the other half the negative spatial correlation, which can be included as explanatory variables in regression models. The scale at which a dbMEM is spatially structured can be estimated as the wavelength of a sine function, visually seen as the distance between peaks of sites with comparable eigenvectors (see Figure 1, for examples, with contrasting spatial structure; Appendix 2 shows additional dbMEMs).

We used dbMEMs [dbmem() function from adespatial], where the spatial weighting matrix is derived from the geographical distance between sites, keeping only the distances that are smaller than the threshold (i.e., distance greater that longest link of the minimum spanning tree, *thresh*, are truncated to 4**thresh*) to build the spatial eigenvectors. We used dbMEMs because we had no a priori ecological expectations for the role of connectivity in the spatial structure across a wide range of climate, land cover and land use and lake morphometric, chemical and biological variables. However, we examined the sensitivity of our results to using dbMEMs as opposed to the more general MEMs by building spatial eigenvectors with different connectivity and geographical distance weighting matrices [using the mem() function from the adespatial package in R]. We found our results to be robust; changing the connectivity and weighing matrices had very little effect on the total amount of variation explained by dbMEMs versus MEMs (Supporting Information). Therefore, our results were not sensitive to the choice of the connectivity and weighting matrices, providing more confidence in our conclusions and our use of dbMEMs.

We then quantified the total strength of spatial structure in each ecological variable and the main scales at which a variable was spatially structured. For each ecological variable, we performed a forward selection regression (i.e., spatial filtering) with all potential dbMEMs (with significant Moran's I) as predictor variables, using vegan's ordiR2step() function, with the direction argument set to "forward" selection (Blanchet, Legendre, & Borcard, 2008). Variables were transformed for normality ($\log_{10}$ transformation for Runoff, Baseflow, Catchment slope, TN, TP, Secchi, Chl a, Lake area, Lake perimeter, Lake depth and Watershed area; logit transformation for land cover/land use variables; no transformation for climate and atmospheric deposition variables). Owing to the high number of data points, many dbMEMs were significant predictors, but few explained meaningful amounts of variation in LAGOS-NE variables; at *c*. 2% of the variance explained, the declining variance explained by each additional dbMEM tended to plateau. Hence, we present only results for dbMEMs that explained ≥ 2% of the variation in LAGOS-NE variables (Table 1).

Given that the goal of our study was to gain a better understanding of how spatial structure varies across a large number of ecosystem and geographical properties and how that spatial structure influences relationships among them, we performed multivariate analyses to group the variables based on spatial similarity (Legendre & Legendre, 2012). In particular, we first conducted a hierarchical agglomerative clustering analysis (Ward's minimum variance method) on the data presented in Table 1 to identify groups of variables that concorded in the importance of dbMEM variables. Here we transposed the matrix presented in Table 1 (dbMEMs as rows, and ecosystem and geographical variables as columns), then calculated the Spearman's r correlation among the ecosystem and geographical variables (columns), which we then further transformed into distances [using the as.dist(1 − absolute(r)) function from the base stats package; R Core Team, 2017], and calculated Ward's agglomerative clustering [using the hclust(…, method = "ward.D2") function from the stats package]. To illustrate the association among the groups of variables, we conducted a principal components analysis on the data presented in Table 1 [ecosystem and geographical variables as rows, and dbMEMs as columns; using the rda() function from the vegan package] and colour coded the ecosystem and geographical variables according to the groups determined by Ward's cluster analysis. We then performed a correspondence analysis (CA) to identify which variables were predominantly structured and at what scale.

Owing to our research questions, we did not spatially detrend the data. Spatial analyses are typically conducted on detrended data in order to account for spatial non-stationarity, and thus for improved ability to capture local patterns that are independent of broad-scale gradients (Fortin & Dale, 2005). However, the objective of our study was to characterize the strength and scales of spatial structure in ecosystem and geographical properties to identify the spatial structure present across scales, both latitudinally and longitudinally. We consider this information meaningful and thus aimed to quantify and explain these potential two-dimensional trends rather than eliminating them. We then qualitatively tested how spatial structure may influence relationships among well-studied variables using the concentrations of

TP, a widely studied ecosystem response in lakes that are subject to cultural eutrophication. We performed a multiple linear regression to identify the main geographical drivers of lake TP across the study extent using the vegan package (Oksanen et al., 2013) in R.

2.3 | Quantifying the role of spatial structure in explaining the strength of relationships between variables

Once we determined the spatial structure for each variable, we quantified the importance of spatial structure for relationships between ecosystem and geographical properties. To do so, we measured the similarity in spatial structure among each possible pair of variables by calculating one minus the difference in the cumulative coefficient of determination (r^2) of each dbMEM–response variable pair. For example, dbMEMs explain 33% of the variation in both Catchment slope and lake TN, hence, they have a similarity of one, whereas mean annual temperature (MAT; cumulative r^2 = 0.92) and watershed area (cumulative r^2 = 0) would have a similarity of 0.08 (Table 1). We then plotted the absolute correlation (i.e., negative values reported as positive), *r*, between pairs of variables (*y* axis) as a function of similarity in spatial structure for the same pair of variables (*x* axis) to explore whether variables with comparable spatial structure tend to have stronger correlations and, if this is the case, the shape and strength of the relationship between correlation and similarity in spatial structure.

There were 264 possible pairs of variables, but 12 pairs were excluded because they contained redundant information (e.g., percentage pasture vs. percentage agricultural land use). Regressions were performed on transformed variables using JMP 10.1 (SAS institute, Cary, NC). We acknowledge that OLS may not be the optimal model for every pair of variables, but here we focus on the overall pattern that emerges across all the pairs of variables rather than on optimizing the variance explained in every pair. Furthermore, we do not attempt to assign a significance value to this analysis of many individual models, and instead focus on the patterns of the relationships among types of variables and their spatial patterns.

Finally, we explored the generality of our results because we acknowledge that the spatial resolution and extent of the study area may impose bias on the scale of the patterns measured (Turner, O'Neill, Gardner, & Milne, 1989; Wu, 2004). Therefore, we

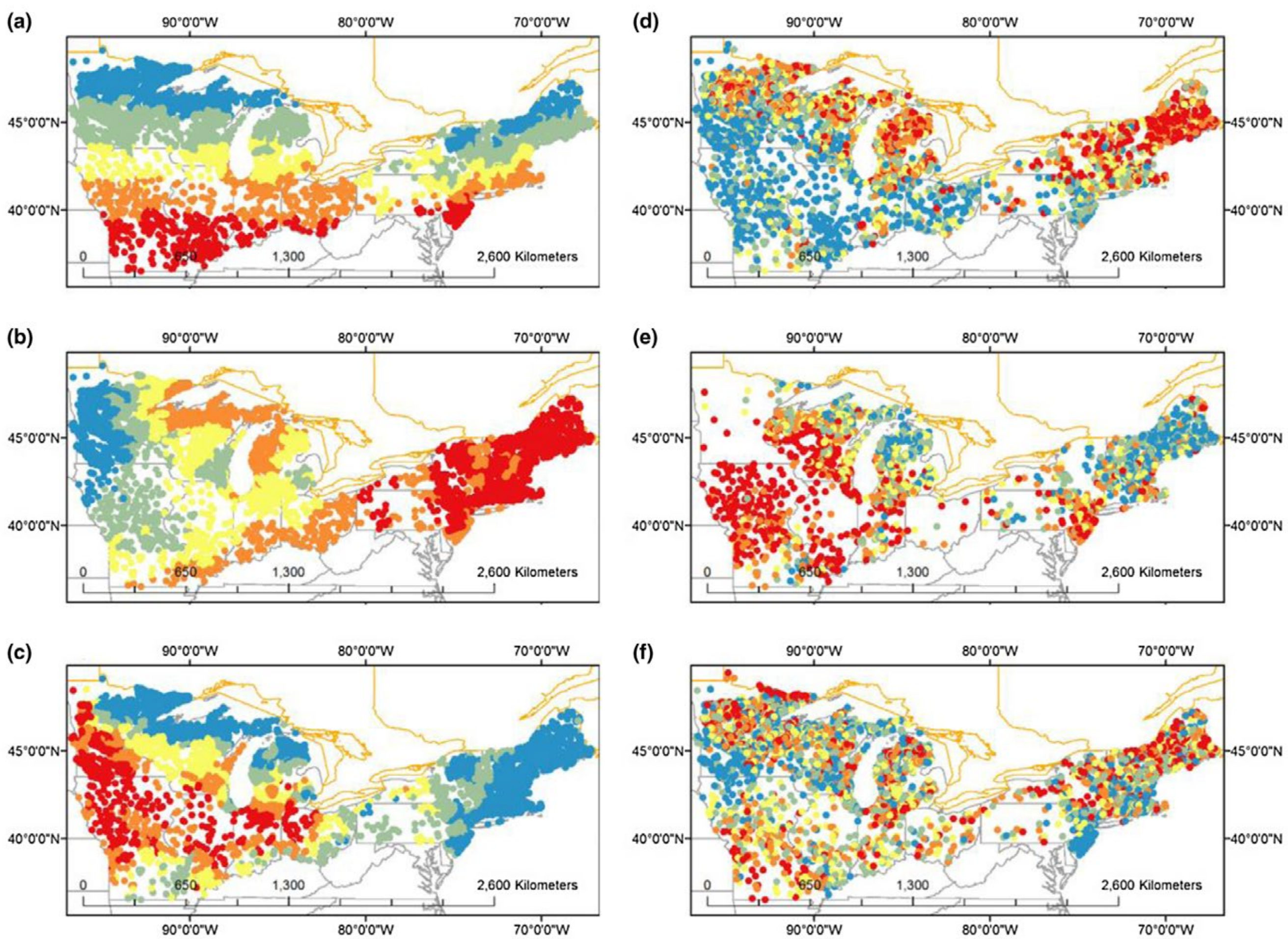

FIGURE 2 Spatial patterns in a subset of ecosystem and geographical properties. Spatial structure decreases from (a) mean annual temperature (MAT), (b) surface runoff, (c) percentage of agriculture in the catchment, (d) mean Secchi depth, (e) mean TP concentrations and (f) maximum lake depth (see Table 1). Values decrease from red to blue, with each colour representing values within a 20 percentile for this variable distribution (range provided in the Methods section) [Colour figure can be viewed at wileyonlinelibrary.com]

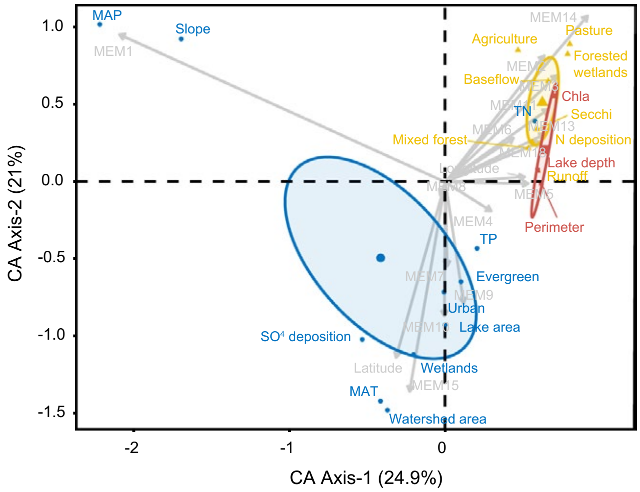

FIGURE 3 Correspondence analysis (CA) and clustering on Table 1. The lengths of the vectors are proportional to the importance of the ecosystem variable to the distance-based Moran eigenvector maps (dbMEMs). The grouping of ecosystem variables was determined by Ward's agglomerative clustering applied on [1 − absolute(*r*)], where confidence ellipses are shown for each [Colour figure can be viewed at wileyonlinelibrary.com]

simulated data that had increasing degrees of spatial autocorrelation using the gstat package (Pebesma & Gr, 2014). We simulated 150 "variables" (i.e., grids of 100 × 100 equally distanced points, unitless) with ranges from 0.01 (near absence of spatial autocorrelation) to 50 (strong spatial autocorrelation). Then, we followed the same steps as for the real dataset. For each simulated variable, we quantified the spatial structure using dbMEMs, then we quantified the similarity in spatial structure and the absolute correlation between each pair of variables. Similar patterns between the LAGOS-NE and simulated variables would suggest mathematical constraints modulating how spatial structure explains the strength of relationships between variables, and that these constraints are independent of scale and of the nature of the input variables.

3 | RESULTS

3.1 | Quantifying spatial structure in ecosystem properties and their geographical context

There were different degrees of spatial structure in climate, landscape and lake properties that were apparent when mapping the quantiles of these variables across the study extent (Figure 2). For example, MAT and, to a lesser degree, annual surface runoff (henceforth, runoff) showed strong linear gradients, generally increasing from North to South and from West to East, respectively. At an intermediate scale, there were strong spatial patterns in the percentage of agricultural land use (Figure 2). Spatial patterns in lake ecosystem properties, such as Secchi and concentrations of TP were also structured at an intermediate scale and appeared to align with agricultural land use patterns, although there are hints of local variation in these lake ecosystem properties that were not apparent for the agricultural land use data (Figure 2). Finally, lake maximum depth (hereafter, depth) showed weak spatial patterns across the study extent (Figure 2).

These observed spatial patterns were validated by the dbMEM analyses. Spatial structure, expressed as the cumulative r^2 between an ecosystem and geographical properties and the most meaningful dbMEMs (see Figure 1, Appendix 2) typically followed a pattern of climate > land cover/land use > lake ecosystem properties > lake and catchment morphometry (Table 1). In particular, dbMEMs collectively explained> 90% of the variation in MAT, precipitation and runoff, indicating very strong spatial structure in these variables at this sub-continental scale, which is well known to be the case for these types of variables. The dbMEMs explained between 30% and 85% of the variation in atmospheric deposition of N and SO_4, various measures of land use/land cover and in catchment slope (Table 1), indicating an intermediate level of spatial structure that is also consistent with what is known about these variables (Figure 2). Interestingly, lake ecosystem properties (concentrations of nutrients, Chl a and Secchi depth; Table 1, in light blue) also had an intermediate spatial structure, comparable to the land cover/land use variables with the weakest spatial structure (Table 1, in green). Finally, dbMEMs explained almost no variation in lake and catchment morphometry (e.g., lake and catchment area, lake depth and perimeter; Table 1, in dark blue), indicating very low spatial structure for these variables. Furthermore, only broad-scale dbMEMs (i.e., the top 10; wavelength of 500–3,200 km) explained meaningful amounts of variation, even in variables that were weakly structured (Table 1).

Beyond the overall strength of spatial structure (i.e., the total amount of variation in ecosystem and geographical variables explained by dbMEMs), there were three main groups of variables with distinct patterns in terms of the spatial scales at which they were predominantly structured. In particular, the spatial structure of variables from group 1 (blue variables in Figure 3) was mainly at a broad scale, as shown by the predominant effect of latitude and dbMEMs 1 (Table 1; Figure 3); for these variables, there tended to be continuous gradients across the study extent. This group includes mainly variables related to climate and atmospheric deposition. Moreover, although the strength of spatial structure in lake area was very low (Table 1), the little spatial structure present tended to be at a broad spatial scale. Variables in group 2 (yellow in Figure 3) are related to land cover/land use and were mostly explained by dbMEM 2, 3 and 14 (Figure 3; Table 1), which had a wavelength of *c*. 200–1,400 km. Finally, spatial structure in variables from group 3 (lake depth, lake perimeter and Chl a; red in Figure 3) were mostly explained by smaller scale dbMEMs, implying that in addition to the overall weaker spatial structure in these variables, the spatial structure present tends to be at a smaller spatial scale.

We then evaluated how spatial structure may influence relationships among ecosystem and geographical properties using concentrations of TP, a widely studied lake ecosystem response variable. Concentrations of TP were best explained in a multiple linear regression model by lake depth, the percentage of evergreen forest in the catchment (%Evergreen), watershed area (WA), runoff and baseflow, in decreasing order of explanatory power (n = 3,110, r^2 = 0.59, p < 0.001):

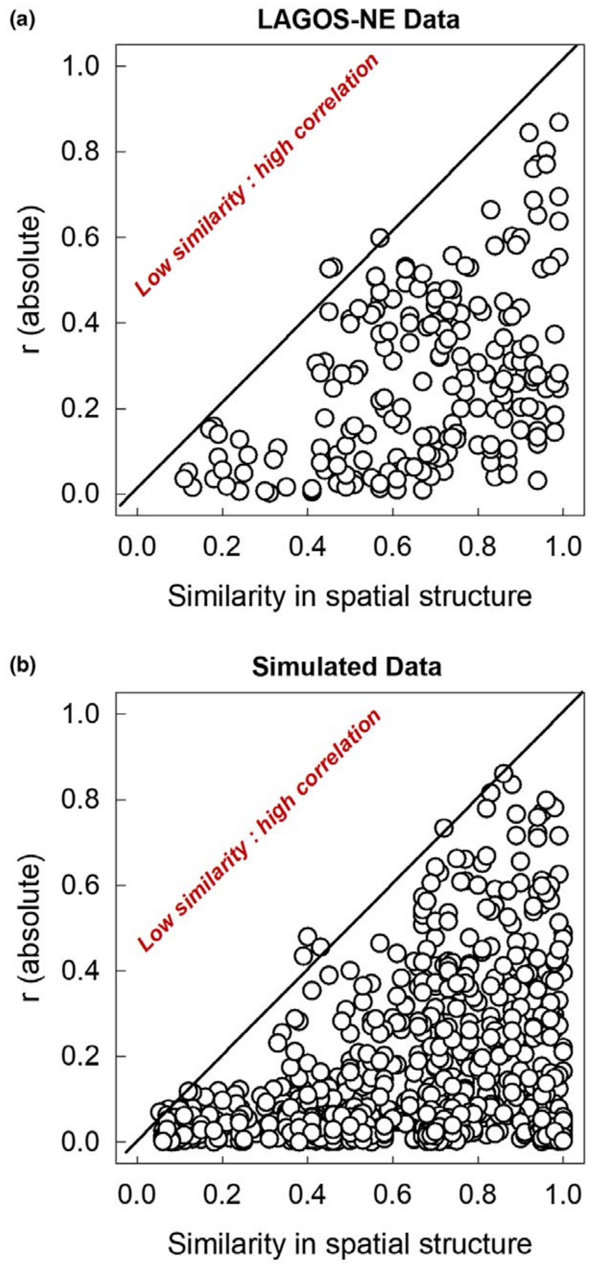

FIGURE 4 The importance of comparable spatial structure in the strength of relationships between pairs of variables. (a) Correlation (absolute values) between pairs of variables is plotted against the similarity in spatial structure (see Table 1) between the corresponding pair of variables in the LAGOS-NE (LAke multi-scaled GeOSpatial and temporal) dataset. (b) The same relationships as in (a), but with a simulated dataset of variables with a known spatial structure (see Methods). Variables are unitless [Colour figure can be viewed at wileyonlinelibrary.com]

$$\begin{aligned} \log(\mathrm{TP}) = & -0.54\log(\mathrm{depth}) - 0.27\mathrm{logit}(\%\mathrm{Evergreen}+1) \\ & +0.12\log(\mathrm{WA}) - 0.015\,(\mathrm{Runoff}) - 0.008\,(\mathrm{Baseflow}). \end{aligned} \quad (1)$$

Spatial structure in lake TP in this study (0.40, expressed as the r^2 between TP and the significant dbMEMs) roughly corresponds to a weighted average of the spatial structure of its main drivers (depth = 0.09, %Evergreen = 0.68, WA = 0, Runoff = 0.90, Baseflow = 0.60; Table 1). Moreover, nearly as much variation in lake TP was explained by purely spatial variables (dbMEMs) containing no inherent ecological information. These results suggest that environmental drivers not only explain variation in TP, but they also induce spatial patterns in it that match that of its main drivers.

3.2 | Similarity in spatial structure and the strength of relationships between variables

The potential strength of relationships between ecosystem and geographical properties increased with similarity in spatial structure. There were 252 possible pairs of variables with different levels of similarity, and no strong relationships were observed when there was strong dissimilarity in spatial structure (Figure 4a). Variables with similar spatial structure, on the contrary, tended to have stronger correlations (Figure 4a). Weak correlations were also observed, however, for variables with similar spatial structure, and the overall pattern showed a triangle shape, where the hypotenuse delineates the maximum potential correlation for a given level of similarity in spatial structure.

We argue that this pattern is generalizable to any type of variable because the results are the same for the simulated dataset (Figure 4b). In particular, simulated variables with known degrees of spatial structure showed an identically shaped pattern, suggesting that this result is not attributable to a bias in the selection of variables or in the spatial extent of our study area. Thus, similar spatial structure appears to be a necessary but not sufficient condition for observing strong relationships between ecosystem and geographical properties such as climate, land cover/land use and indicators of the flow of matter and energy in lakes.

4 | DISCUSSION

The spatial structure of ecosystem and geographical properties contains rich information that can provide insight on the potential drivers and processes that give rise to spatial patterns. Although our approach to quantify spatial structure and its effect on paired relationships is correlative, it illustrates how spatial structure can constrain these types of broad-scale relationships to explain ecosystem variation at macroscales. Understanding these constraints might help to identify unmeasured processes operating at similar spatial scales that affect ecosystem properties. Furthermore, a priori knowledge of spatial structure in ecosystem properties and their potential drivers may allow ecologists to focus hypotheses and sampling efforts on the dominant variables that explain variation in ecosystem properties across scales, whether relationships are directly causal or reflect other mechanisms. Our study shows that inferences can be made about regional to global processes by simply measuring spatial patterns and their effects on correlations between a wide range of ecosystem and geographical properties. We expect our results to apply beyond lake ecosystems and hope that similar analyses will be conducted on other multi-themed datasets at macroscales.

Our interpretations focus on generalities rather than on the details of each possible relationship. There may be biases in the relationships between individual variables that have been measured or averaged at different spatial and temporal grains. For example, we quantified spatial structure in 10-year medians (with varying numbers of observations) for point measurements of lake ecosystem properties, and in 30-year average mean annual precipitation or temperature (MAP and MAT, respectively) at the HU12 level (median size = 78 km^2) that have been interpolated from broadly distributed meteorological stations. Detailed understanding of the potential causal relationships between each pair of variables would necessitate a finer exploration of the scales at which variables have been measured and averaged, but this was not the aim of our study. Rather, the strength of our results lies in the overall pattern that is shown for ecosystem properties reported for> 8,000 sites across a 1.8 million km^2 area. Moreover, the fact that we can reproduce our key pattern with simulated variables that are unitless and unbiased (Figure 4b) suggests that the patterns are robust and generalizable beyond the study extent and the variables studied here. In particular, although we focus on climate, landscape and lake ecosystem properties over a broad study extent, our simulations suggest that our results are applicable across study extent and for different types of ecological, ecosystem and geographical data.

4.1 | Similarity in spatial structure as a necessary condition for correlations between ecosystem properties and drivers

The connection between spatial structure and correlations between pairs of variables is best illustrated in two zones of Figure 4. The first zone, in the upper-left triangle, contained no pairs of variables for either of the datasets (measured or simulated), suggesting that there cannot be strong correlations between ecosystem properties with very different spatial structures. From a predictive point of view, this means that to predict ecosystem properties at broad spatial extents, one should focus on measuring drivers that have broad-scale spatial structure. This result also means that if variation in an ecosystem property was meaningfully explained by a driver variable with a given spatial structure, that variable would induce some of its spatial structure in the response variable, and thus this pair of variables would be found in the lower-right triangle of Figure 4. Does this mean that no pairs of variables can be found in the "forbidden zone" (top-left triangle)? Teleconnections that decouple the spatial scales of an ecosystem property and its immediate drivers (e.g., trade, migrations; Heffernan et al., 2014; Pace & Gephart, 2017) could, perhaps, generate pairs of variables fitting in this region, but to the best of our knowledge there are no empirical tests of these relationships.

The second important zone of Figure 4 is the bottom-right triangle, which shows the realm of potential relationships, where the maximum correlation gradually increases with similarity in spatial structure. Pairs of variables in the left part of this zone include lake and watershed morphometry (e.g., area) with long-term climate averages (e.g., MAP), which have the most different spatial structure (Table 1) and are never strongly correlated (Figure 4). For example, if climate had any tangible effect on lake or watershed area (or vice versa), it would induce some degree of spatial structure, resulting in comparable spatial patterns. Although correlations are not sufficient to demonstrate causation, correlations imply an unresolved causal structure, whereby the effect can be indirect and explained by some unmeasured variable (Shipley, 2002). Therefore, the near absence of correlations in this part of the figure suggests that variables with completely different spatial structure do not have meaningful effects on each other. Interestingly, however, even the weak spatial structure exhibited by lake morphometry variables (e.g., lake depth, lake area) was explained by broad-scale dbMEM (Table 1; Figure 3), similar to their weak but significant correlations with climate variables. Although variables, such as climate and lake area may not be related causally at the present time [although Downing et al. (2006) observed higher occurrence of farm ponds where MAP > 1,600 mm/year], historical climate might have set a broader sub-continental context. For example, the Late Wisconsinian glaciation divides our study area into glaciated and unglaciated regions, leading to lasting effects on landscape configuration (Fergus et al., 2017), even though the effects of glaciation have been muted over time.

Pairs of variables in the upper-right part of this zone include variables with similar spatial structure and strong relationships, such as climate variables (MAT or MAP) versus atmospheric deposition (N or SO_4) variables, runoff versus MAP and lake area versus lake depth and perimeter. In addition to meeting the necessary, mathematical constraint of having similar spatial structure, these variables share well-known causative links. For example, it is expected that runoff will be high in areas with high MAP because more water will be received in these landscapes, and large lakes are likely to have a large perimeter and to be deep. Likewise, the specific relationship between lake area and lake perimeter is confirmed by well-known geometric constraints (Cael & Seekell, 2016; Cael, Heathcote, & Seekell, 2017; Seekell, Pace, Tranvik, & Verpoorter, 2013). Thus, some of the pattern in this figure is attributable to geometric or physical constraints, or from closely correlated variables that are tied to a nearly identical process (i.e., rainfall). However, the results from the simulated data provide strong evidence that within a given spatial extent for a study, the spatial pattern of the variables themselves impose this same basic pattern, regardless of these geometric constraints or common processes.

Finally, the lower-right part of this second zone includes variables with similar spatial structure that are not strongly correlated. These relationships suggest that although having similar spatial structure is a necessary condition for having strong correlations between pairs of variables, it is not a sufficient one. For example, this zone is where relationships exist between lake ecosystem properties with intermediate, regional spatial structure (e.g., lake TN, TP, Chl a, Secchi) and land cover/land use variables (that also exhibit intermediate, regional spatial structure). Given that ecosystem properties are likely to be explained by several, potentially interacting drivers operating at multiple spatial scales, causation is not as obvious as it is for relationships found in the top right portion of Figure 4, nor is it a simple function of the similarity in spatial structure. As a result, pairs of variables are not as strongly correlated even though they have a similar spatial structure, either because there are local, intrinsic effects (e.g., biotic processes) that decouple patterns from their broader physical constraints (Fortin & Dale, 2005; Levin, 1992) or because variables randomly happen to have a similar spatial structure but do not share any causative links. This is where multiple effects are confounded, thus where finer scale (spatial or temporal) ecological interpretation is required to understand ecosystem functioning.

4.2 | Considering spatial structure for a better understanding and prediction of ecosystem relationships

There has been much written about how the spatial extent and grain of studies influence the understanding and prediction of organisms and ecological properties (Horne & Schneider, 1995; O'Neill, DeAngelis, Waide, & Allen, 1986; Wiens, 1989). The manner in which drivers act together and their relative importance at differing spatial scales determines species distributions and biological spatial structuring (McGill, 2010; Russell, Wood, Allison, & Menge, 2006). In general, species interactions in a variety of taxa, including avian guilds, intertidal alga and invertebrates, have been found to be important at small to mid-size spatial scales but are unimportant at the scale of biomes or continents (Gotelli et al., 2010; McGill, 2010; Russell & Connell, 2012; Veech, 2006). Environmental filtering of biological communities, owing to constraints imposed by broad-scale climate and habitat characteristics, is prevalent at mid- to large spatial scales in the distribution of biological pathogens and many other taxa, but such drivers may be disrupted at smaller scales by species interactions and fine-scale weather or habitat features (Cohen et al., 2016; Wiens, 1989). Furthermore, evidence from studies of food chain length in lakes has been found to be correlated with lake volume for small numbers of lakes within a limited geographical area but not for lakes at the global scale (Post, Pace, & Hairston, 2000; Vander Zanden & Fetzer, 2007). Findings from biological communities thus suggest that spatial structure in ecological properties is controlled by a composite of multiple interacting drivers that differ depending on spatial extent.

Our empirical findings from thousands of ecosystems at the macroscale suggest that these principles are transposable to understand the drivers of spatial structure of matter and energy in ecosystems. For example, the spatial structure of an ecosystem property should be a composite of the spatial structure of its main drivers, and we showed that concentrations of TP in 3,452 lakes were explained by a suite of well-known drivers of TP (Equation 1). Therefore, full interpretation of this simple model needs to account for known ecological relationships and the inherent spatial structure in the variables involved (Table 1; Figure 4). It appears that broad-scale patterns in hydrology and land cover at regional and continental scales induce strong spatial structure in lake TP across the study extent, but that lake and catchment morphometry, which have been associated with terrestrial loadings and in-lake processing of nutrients (Collins et al., 2017; Read et al., 2015), affect lake TP locally and tend to mute the broad-scale effects (see Figure 2e). Along the same line, patterns in lake dissolved organic carbon (DOC) within an individual region are mainly related to lake area and perimeter (Frost et al., 2006), but at the global scale the main drivers are long-term averages of precipitation, runoff and soil carbon content (Sobek, Tranvik, Prairie, Kortelainen, & Cole, 2007). Other studies have shown that carbon and nitrogen cycles in high-latitude catchments have been linked to climate and permafrost at continental scales, but to watershed characteristics at regional scales (Harms et al., 2016). Together, these studies show the scale dependence of ecosystem-level relationships in different regions of the globe, and here we have quantified this effect and explicitly demonstrated the importance of spatial structure in understanding ecosystem relationships at the macroscale.

In addition to the underlying ecological mechanisms driving spatial patterns in ecosystem properties, our results further suggest that there are mathematical constraints affecting the statistical relationships between variables based on their respective spatial structure, even in the absence of meaningful ecological links. A key finding of our study is that correlation is stronger for variables with similar spatial structure, and that the strength of the correlation gradually decreases with dissimilarity in spatial structure, for both measured and simulated variables. We visually expressed the importance of considering the spatial structure of variables by rearranging the points from Figure 4a; we first extracted the variables with weak, intermediate and strong cumulative r^2 with the dbMEMs (Table 1; cumulative r^2 ranging from 0 to 0.3, 0.31 to 0.70 and 0.71 to 0.99, respectively). For each of these groups, we then plotted their respective correlation with all the other variables, as a function of strength in spatial structure. Weakly structured variables tended to be most strongly correlated with other weakly structured variables, and the strength of correlation gradually decreased with variables with intermediate and strong spatial structure (Figure 5a). Likewise, variables with intermediate spatial structure tended to be most strongly correlated with other variables having intermediate spatial structure (Figure 5b), and strongly structured variables tended to be most strongly correlated with variables with strong spatial structure (Figure 5c). Figure 5d shows the basic patterns in these variables for ecosystem and geographical properties that should apply to a wide range of lake properties and, presumably, other types of ecosystems. These results imply that studies conducted over small spatial extents are much more likely to find that locally structured predictor variables explain the most variation in response variables (e.g., lake and watershed morphometry), whereas continental to global scale studies are more likely to explain meaningful amounts of variation in their response variables with predictor variables that have very broad-scale spatial structure (e.g., atmospheric deposition, long-term climate average). This, in turn, is likely to explain why studies conducted at different spatial extents have often identified different variables as the main drivers of ecosystem processes.

Based on an analysis of >8,000 sites in a sub-continental study area of 1.8 million km^2, our study highlights a need to understand spatial structure for improved understanding of the relationships between climate, landscape and ecosystem properties. Ecosystem studies are

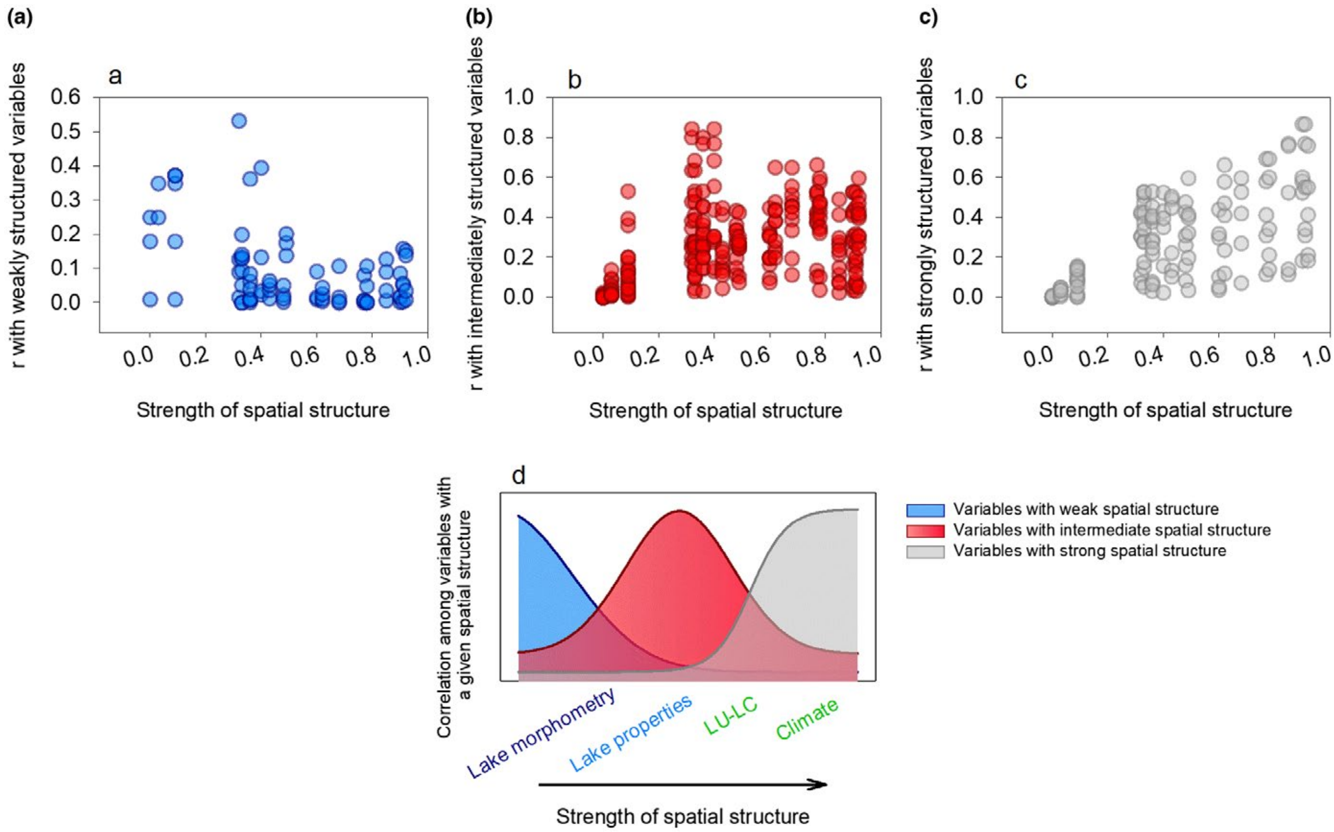

FIGURE 5 The relationship between spatial structure and correlations among pairs of variables. (a) Variables with weak spatial structure tend to be more strongly related to other variables that have weak spatial structure. (b) Variables with intermediate spatial structure tend to be most strongly correlated with variables of intermediate to high spatial structure. (c) Variables with strong spatial structure tend to be most highly correlated with other variables with intermediate to strong spatial structure. (d) The strength of spatial structure increases from lake morphometry variables to climate variables, with lake ecological properties (lake properties) and land use/land cover (LU-LC) variables having intermediate spatial structure (colour codes for "*x*" labels match these of Table 1). Lake ecosystem properties tend to have the strongest relationships with comparably structured land use/land cover variables at broad spatial scales [Colour figure can be viewed at wileyonlinelibrary.com]

increasingly conducted at broad spatial scales, with the associated conceptual and analytical challenges. Here, we provide a conceptual basis and suggest approaches to consider explicitly a spatial aspect in ecosystem relationships, with the aim of improving the understanding and prediction of ecosystem relationships from local to global scales.

ACKNOWLEDGMENTS

This research was supported by the NSF Macrosystems Biology program (awards EF-1065786 and EF-1065818), a postdoctoral fellowship from the Fonds de Recherche du Québec en Nature et Technologies and a Discovery Grant from the National Science and Engineering Research Council of Canada to J.-F.L., a United States Department of Agriculture (USDA) National Institute of Food and Agriculture, Hatch Project grant to P.A.S. (no. 176820), an National Science Foundation (NSF) Postdoctoral Research Fellowship in Biology to S.M.C. (DBI-1401954), and the Knut and Alice Wallenberg Foundation. We thank the Cross-scale interactions-Limnology research group and P. Legendre for constructive discussions during the initial and revision stages of this manuscript.

DATA ACCESSIBILITY

The data used in this manuscript are available at: https://dx.doi.org/10.6073/pasta/4e479018aa3d23603736bb5849f9a200. The R code used is published as Supporting Information Appendix S1.

ORCID

Jean-Francois Lapierre http://orcid.org/0000-0001-5862-7955

REFERENCES

Anselin, L. (2001). Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics*, *83*, 705–710. https://doi.org/10.1111/0002-9092.00194

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

Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth's lakes. *Geophysical Research Letters*, *44*, 209–218. https://doi.org/10.1002/2016GL071378

Cael, B. B., & Seekell, D. A. (2016). The size-distribution of Earth's lakes. *Scientific Reports*, *6*, 29633. https://doi.org/10.1038/srep29633

Cheruvelil, K., Soranno, P., Bremigan, M., Wagner, T., & Martin, S. (2008). Grouping lakes for water quality assessment and monitoring: The roles of regionalization and spatial scale. *Environmental Management*, *41*, 425–440. https://doi.org/10.1007/s00267-007-9045-7

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., … Wear, D. (2001). Ecological forecasts: An emerging imperative. *Science*, *293*, 657–660. https://doi.org/10.1126/science.293.5530.657

Cohen, J. M., Civitello, D. J., Brace, A. J., Feichtinger, E. M., Ortega, C. N., Richardson, J. C., … Rohr, J. R. (2016). Spatial scale modulates the strength of ecological processes driving disease distributions. *Proceedings of the National Academy of Sciences USA*, *113*, E3359–E3364. https://doi.org/10.1073/pnas.1521657113

Collins, S. M., Oliver, S. K., Lapierre, J. F., Stanley, E. H., Jones, J. R., Wagner, T., & Soranno, P. A. (2017). Lake nutrient stoichiometry is less predictable than nutrient concentrations at regional and subcontinental scales. *Ecological Applications*, *27*, 1529–1540. https://doi.org/10.1002/eap.1545

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., … Middelburg, J. J. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, *51*, 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Dray, S., Blanchet, F. G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., … Wagner, H. H. (2016). *Adespatial: Multivariate multiscale spatial analysis*. (R package version 0.0-3).

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*, 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Fergus, C. E., Lapierre, J. F., Oliver, S. K., Skaff, N. K., Cheruvelil, K. S., Webster, K., … Soranno, P. (2017). The freshwater landscape: Lake, wetland, and stream abundance and connectivity at macroscales. *Ecosphere*, *8*, e01911. https://doi.org/10.1002/ecs2.1911

Fortin, M.-J., & Dale, M. R. T. (2005). *Spatial analysis: A guide for ecologists*. Cambridge: Cambridge University Press.

Frost, P. C., Larson, J. H., Johnston, C. A., Young, K. C., Maurice, P. A., Lamberti, G. A., & Bridgham, S. D. (2006). Landscape predictors of stream dissolved organic matter concentration and physicochemistry in a Lake Superior river watershed. *Aquatic Sciences*, *68*, 40–51. https://doi.org/10.1007/s00027-005-0802-5

Gotelli, N. J., Graves, G. R., & Rahbek, C. (2010). Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences USA*, *107*, 5030–5035. https://doi.org/10.1073/pnas.0914089107

Harms, T. K., Edmonds, J. W., Genet, H., Creed, I. F., Aldred, D., Balser, A., & Jones, J. J. (2016). Catchment influence on nitrate and dissolved organic matter in Alaskan streams across a latitudinal gradient. *Journal of Geophysical Research: Biogeosciences*, *121*, 350–369. https://doi.org/10.1002/2015JG003201

Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., … Weathers, K. C. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, *12*, 5–14. https://doi.org/10.1890/130017

Hoeting, J. A. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, *19*, 574–577. https://doi.org/10.1890/08-0836.1

Horne, J. K., & Schneider, D. C. (1995). Spatial variance in ecology. *Oikos*, *74*, 18–26. https://doi.org/10.2307/3545670

Lapierre, J.-F., Collins, S., Seekell, D., Cheruvelil, K., Tan, P., Skaff, N., … Soranno, P. (2017). *LAGOS-NE—Lake nutrient chemistry and geospatial data to measure spatial structure of ecosystem properties in a 17-state region of the U.S. Environmental Data Initiative*. https://doi.org/10.6073/pasta/4e479018aa3d23603736bb5849f9a200. Dataset accessed 7/18/2018.

Lapierre, J.-F., Seekell, D. A., & del Giorgio, P. A. (2015). Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. *Global Change Biology*, *21*, 4425–4435. https://doi.org/10.1111/gcb.13031

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*, 1659–1673. https://doi.org/10.2307/1939924

Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed., p. 1006). Amsterdam, the Netherlands: Elsevier.

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

Levy, O., Ball, B. A., Bond-Lamberty, B., Cheruvelil, K. S., Finley, A. O., Lottig, N. R., … Williams, J. W. (2014). Approaches to advance scientific understanding of macrosystems ecology. *Frontiers in Ecology and the Environment*, *12*, 15–23. https://doi.org/10.1890/130019

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., … Pinay, G. (2003). Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems*, *6*, 301–312. https://doi.org/10.1007/s10021-003-0161-9

McGill, B. J. (2010). Matters of scale. *Science*, *328*, 575–576. https://doi.org/10.1126/science.1188528

McGuire, K. J., Torgersen, C. E., Likens, G. E., Buso, D. C., Lowe, W. H., & Bailey, S. W. (2014). Network analysis reveals multiscale controls on stream water chemistry. *Proceedings of the National Academy of Sciences USA*, *111*, 7030–7035. https://doi.org/10.1073/pnas.1404820111

O'Neill, R. V., DeAngelis, D. L., Waide, J. B., & Allen, T. F. H. (1986). *A hierarchical concept of ecosystems*. Princeton, NJ: Princeton University Press.

O'Reilly, C. M., Rowley, R. J., Schneider, P., Lenters, J. D., McIntyre, P. B., & Kraemer, B. M. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, *42*, 10773–10781. https://doi.org/10.1002/2015GL066235

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2013). *Vegan: Community ecology package*.

Pace, M. L., & Gephart, J. A. (2017). Trade: A driver of present and future ecosystems. *Ecosystems*, *20*, 44–53. https://doi.org/10.1007/s10021-016-0021-z

Pebesma, E., & Gräler, B. (2013). Spatio-temporal geostatistics using gstat. Institute for Geoinformatics, University of Münster Rep.

Post, D. M., Pace, M. L., & Hairston, N. G. (2000). Ecosystem size determines food-chain length in lakes. *Nature*, *405*, 1047–1049. https://doi.org/10.1038/35016565

Qiu, J., & Turner, M. G. (2013). Spatial interactions among ecosystem services in an urbanizing agricultural watershed. *Proceedings of the National Academy of Sciences USA*, *110*, 12149–12154. https://doi.org/10.1073/pnas.1310539110

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, *8*, 224–239. https://doi.org/10.1111/j.1461-0248.2004.00701.x

Read, E. K., Patil, V. P., Oliver, S. K., Hetherington, A. L., Brentrup, J. A., Zwart, J. A., … Weathers, K. C. (2015). The importance of lake-specific characteristics for water quality across the continental United States. *Ecological Applications*, *25*, 943–955. https://doi.org/10.1890/14-0935.1

Rose, K. C., Graves, R. A., Hansen, W. D., Harvey, B. J., Qiu, J., Wood, S. A., … Turner, M. G. (2017). Historical foundations and future directions in macrosystems ecology. *Ecology Letters*, *20*, 147–157. https://doi.org/10.1111/ele.12717

Russell, B. D., & Connell, S. D. (2012). Origins and consequences of global and local stressors: Incorporating climatic and non-climatic phenomena that buffer or accelerate ecological change. *Marine Biology*, *159*, 2633–2639. https://doi.org/10.1007/s00227-011-1863-8

Russell, R., Wood, S. A., Allison, G., & Menge, B. A. (2006). Scale, environment, and trophic status: The context dependency of community saturation in rocky intertidal communities. *The American Naturalist*, *167*, E158–E170. https://doi.org/10.1086/504603

Seekell, D. A., Lapierre, J. F., Pace, M., Gudasz, C., Sobek, S., & Tranvik, L. (2014). Regional-scale variation of dissolved organic carbon concentrations in Swedish lakes. *Limnology and Oceanography*, *59*, 1612–1620. https://doi.org/10.4319/lo.2014.59.5.1612

Seekell, D. A., Pace, M. L., Tranvik, L. J., & Verpoorter, C. (2013). A fractal-based approach to lake size-distributions. *Geophysical Research Letters*, *40*, 517–521. https://doi.org/10.1002/grl.50139

Sharma, S., Gray, D. K., Read, J. S., O'Reilly, C. M., Schneider, P., Qudrat, A., … Woo, K. H. (2015). A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009. *Scientific Data*, *2*, 150008. https://doi.org/10.1038/sdata.2015.8

Shipley, B. (2002). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference* (2nd ed.). Cambridge: Cambridge University Press.

Sobek, S., Tranvik, L. J., Prairie, Y. T., Kortelainen, P., & Cole, J. J. (2007). Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. *Limnology and Oceanography*, *52*, 1208–1219. https://doi.org/10.4319/lo.2007.52.3.1208

Soranno, P. A., Bacon, L. C., Beauchene, M., Bednar, K. E., Bissell, E. G., Boudreau, C. K., … Yuan, S. (2017). LAGOS-NE: A multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. lakes. *Gigascience*, *6*, 1–22. https://doi.org/10.1093/gigascience/gix101

Soranno, P. A., Bissell, E. G., Cheruvelil, K. S., Christel, S. T., Collins, S. M., Fergus, C. E., … Webster, K. E. (2015). Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science and data reuse. *GigaScience*, *4*, 28. https://doi.org/10.1186/s13742-015-0067-4

Soranno, P. A., Cheruvelil, K. S., Bissell, E. G., Bremigan, M. T., Downing, J. A., Fergus, C. E., … Webster, K. E. (2014). Cross-scale interactions: Quantifying multi-scaled cause–effect relationships in macrosystems. *Frontiers in Ecology and the Environment*, *12*, 65–73. https://doi.org/10.1890/120366

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: Recent progress and future challenges. *Journal of the North American Benthological Society*, *29*, 344–358. https://doi.org/10.1899/08-171.1

Turner, M. G., O'Neill, R. V., Gardner, R. H., & Milne, B. T. (1989). Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology*, *3*, 153–162. https://doi.org/10.1007/BF00131534

Vander Zanden, J. M., & Fetzer, W. (2007). Global patterns of aquatic food chain length. *Oikos*, *116*, 1378–1388. https://doi.org/10.1111/j.0030-1299.2007.16036.x

Veech, J. A. (2006). A probability-based analysis of temporal and spatial co-occurrence in grassland birds. *Journal of Biogeography*, *33*, 2145–2153. https://doi.org/10.1111/j.1365-2699.2006.01571.x

Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology*, *3*, 385. https://doi.org/10.2307/2389612

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*, 125–138. https://doi.org/10.1023/B:LAND.0000021711.40074.ae

BIOSKETCH

We are an interdisciplinary research group developing and using approaches from a wide range of fields, including limnology, landscape limnology, geographical information science, ecoinformatics, data mining, machine learning and statistics. Our common aim is to create the needed infrastructure of people, data and computer tools for studying water quality at broad scales for research, management, policy, education and outreach. Our approach is based strongly on a foundation of open science, team science and data-intensive science. We believe that by taking an open-science perspective and combining site-based lake datasets and national geospatial datasets, scientists gain the ability to ask important research questions related to grand environmental challenges that operate at macroscales. More information on our research group can be found at: https://lagoslakes.org/about/

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lapierre J-F, Collins SM, Seekell DA, et al. Similarity in spatial structure constrains ecosystem relationships: Building a macroscale understanding of lakes. *Global Ecol Biogeogr*. 2018;27:1251–1263. https://doi.org/10.1111/geb.12781