esa ECOSPHERE

The freshwater landscape: lake, wetland, and stream abundance and connectivity at macroscales

Carol Emi Fergus,[1,8,]† Jean-François Lapierre,[3,4] Samantha K. Oliver,[5] Nicholas K. Skaff,[1] Kendra S. Cheruvelil,[1,6] Katherine Webster,[1,7] Caren Scott,[2] and Patricia Soranno[1]

[1]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824 USA
[2]NEON, 1685, 38th Street, Suite 100, Boulder, Colorado 80301 USA
[3]Département de Sciences Biologiques, Université de Montréal, Pavillon Marie-Victorin, 90, Avenue Vincent-d'Indy, Montreal, Québec H2V 2S9 Canada
[4]GRIL Université de Montréal, Pavillon Marie-Victorin, Département de Sciences Biologiques, C.P. 6128, succ. Centre-ville, Montreal, Québec H3C 3J7 Canada
[5]Center for Limnology, University of Wisconsin Madison, Madison, Wisconsin 53706 USA
[6]Lyman Briggs College, Michigan State University, East Lansing, Michigan 48825 USA
[7]School of Natural Sciences, Trinity College Dublin, The University of Dublin, College Green, Dublin 2, Ireland

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

Abstract. Inland water bodies and their surface hydrologic connections are active components in the landscape, influencing multiple ecological processes that can propagate to broad-scale phenomena such as regional nutrient and carbon cycles and metapopulation dynamics. However, while lake, wetland, and stream abundance has been estimated at regional and global extents, less attention has been paid to freshwater connectivity attributes among aquatic systems at macroscales. Thus, regional to continental patterns of freshwater abundance and connectivity are poorly understood. We measured lake, wetland, and stream abundance and surface connectivity attributes (i.e., landscape position within stream networks) at a subcontinental extent in the Midwest and Northeast United States to characterize macroscale spatial patterns of the freshwater landscape (i.e., abundance and connectivity attributes of lakes, wetlands, and streams). We found that lake and wetland abundance exhibited opposite spatial patterns from stream density that generally followed glaciation extent boundaries—lake and wetland abundance was high north of the glaciation boundary, whereas stream density was high south of the glaciation boundary. Freshwater connectivity attributes exhibited distinct spatial patterns as defined by our integrated freshwater clusters and revealed a layer of complexity not captured by abundance measures. Patterns of freshwater abundance and connectivity in the study extent were associated primarily with glaciation and secondarily with hydrogeomorphic (e.g., surficial geology and topography), climate (e.g., runoff), and land-use (e.g., agriculture) variables, providing insight into potential drivers of freshwater composition and distribution. The connectivity spatial patterns observed suggest that relying solely on freshwater abundance measures in macroscale analyses omits unique information on the structural attributes of freshwater systems that can be critical to key ecological processes. Adopting an integrated freshwater landscape framework to study and manage freshwaters is essential as freshwater systems face broad-scale disturbances that may alter hydrologic connections and subsequently may impact ecosystem processes and services.

Key words: cluster analysis; freshwater abundance; freshwater hydrologic connectivity; glaciation; integrated freshwater landscape; lakes; macroscale; streams; wetlands.

Received 13 March 2017; revised 29 June 2017; accepted 3 July 2017. Corresponding Editor: Debra P. C. Peters.

[8] Present address: The National Research Council, Environmental Protection Agency, Corvallis, Oregon 97333 USA.
† **E-mail:** emifergus@gmail.com

Introduction

Freshwater ecosystems are active components in the landscape that influence a number of ecological processes and provide invaluable ecosystem services. Freshwater connections that link lakes, wetlands, and streams to one another and the surrounding landscape are major structural components that influence critical freshwater processes such as water, nutrient, and carbon fluxes (Quinlan et al. 2003, Cardille et al. 2007, Acuña and Tockner 2010, Lottig et al. 2011), nutrient and carbon processing (Weller et al. 1996, Kling et al. 2000, Strayer et al. 2003), and the dispersal of organisms influencing population and community dynamics (Pringle 2001, Crump et al. 2007, Bouvier et al. 2009, Nelson et al. 2009). These processes can influence broad-scale ecological phenomena with implications for regional and global biogeochemical budgets and metapopulation dynamics. Quantifying patterns of freshwater abundance and connectivity is essential to inform upscaling of biogeochemical processing and for conservation and management of populations at broad spatial extents. However, we have an incomplete view of the integrated freshwater landscape that includes lakes, wetlands, and streams and their spatial connections to one another at macroscales, defined as regional to continental extents spanning hundreds to thousands of kilometers (Heffernan et al. 2014). While progress has been made on quantifying the abundance of individual ecosystem types across the landscape, an integrative picture of the combined freshwater landscape that explicitly includes spatial connections among these systems has not been developed.

Regional and global abundance has been estimated for lakes (Downing et al. 2006, McDonald et al. 2012, Verpoorter et al. 2014), wetlands (Aselmann and Crutzen 1989, Lehner and Döll 2004), and streams (Downing et al. 2012). These studies were the first to explicitly estimate the distribution of freshwater systems across broad spatial extents and examine underlying geospatial features related to these patterns (i.e., glacial and tectonic activity, geomorphology, and regional climate; Meybeck 1995, Smith et al. 2002, Downing and Duarte 2009). Most of these studies, however, were restricted to one freshwater system type, thus providing a limited view of the diversity of the freshwater landscape. Exceptions to this were studies on macroscale carbon cycles that integrated inland waters by estimating surface area of lakes, reservoirs, and streams (Raymond et al. 2013, Butman et al. 2016). However, both single- and integrated-system studies did not distinguish freshwater systems by their hydrologic connectivity attributes (e.g., drainage, seepage system types), and thus provide general abundance estimates that lump all system types together, ignoring the diversity of hydrologic types across the landscape.

Unlike abundance patterns, freshwater connectivity patterns have not been quantified at regional to global scales, in part because freshwater connectivity can be challenging to define (Ali and Roy 2009), is scale dependent, and fine-scale data on hydrologic flow paths are lacking at macroscale extents. Hence, it is challenging to get precise measures of hydrologic connections among systems at broad spatial extents. However, landscape metrics are useful tools that can address this limitation for broad-scale analyses by capturing aspects of freshwater connectivity within the landscape, which we define as the surface connections linking lakes, wetlands, and streams. Landscape position of freshwater systems is one useful class of metrics that is inherently associated with differences in connectivity characteristics moving along longitudinal gradients of geomorphic and hydrologic settings (Kratz et al. 1997, Soranno et al. 1999, Riera et al. 2000, Martin and Soranno 2006, Müller et al. 2013). For example, lake hydrologic position metrics incorporate lake connections with stream inflows and outflows (Martin and Soranno 2006), and lake order differentiates systems based on their position within stream networks (Riera et al. 2000). While landscape position measures do not provide explicit measures of all possible hydrologic flow paths, they do group systems with similar spatial and landscape characteristics, which can be useful to characterize surface connectivity attributes.

Lake, wetland, and stream landscape position metrics are related to biogeochemical variables (Lohse et al. 2009, Humborg et al. 2010, Racchetti et al. 2010, Sadro et al. 2011), hydrogeomorphic and limnological characteristics (Martin and Soranno 2006, Butman and Raymond 2011, Read et al. 2015), and responses to land-use disturbances

(Detenbeck et al. 1993, Freeman et al. 2007, Soranno et al. 2015*a*), demonstrating that they capture spatial connectivity characteristics that are relevant to ecological processes. For example, within a hydrologic flow system, lakes at high positions tend to be hydrologically dominated by precipitation and, thus, have lower base cation concentrations compared to lakes at lower positions that have greater contributions from groundwater and surface water sources to their water budgets (Kratz et al. 1997, Soranno et al. 1999). These landscape position measures are related to ecosystem structure (i.e., surface connectivity attributes) and function and can provide insight on populations of freshwater systems at macroscales.

There is a need to examine freshwater connectivity attributes at broad scales because they are vulnerable to disturbances such as land-use conversion and climate change that can alter hydrologic properties at multiple spatial and temporal scales and subsequently threaten freshwater integrity and function in the landscape (Carpenter et al. 2011, Steele and Heffernan 2014). The effects of broad-scale disturbances on freshwater systems are likely to vary across the landscape depending on underlying hydrogeomorphic characteristics (Webster et al. 2008), the spatial configuration of freshwater features (Vörösmarty et al. 2010), and interactions with natural features and human hydrological modifications (Jones et al. 2012). Thus, it is challenging to predict the impacts of these potential disturbances without understanding freshwater systems within the context of their geographic setting and at the broad spatial scales that are aligned with the above scales of disturbance (Jones 2011, Moore 2015). Together, these issues create a need to develop broad, macroscale estimates of the abundance and connectivity of freshwater systems in the landscape.

Until recently, it has been challenging to incorporate freshwater connectivity characteristics into broad-scale frameworks because of computational limitations and a lack of integrated lake, wetland, and stream datasets. Technological advances and national-scale geographic data resources (e.g., the high-resolution U.S. National Hydrography Dataset; NHD) have ameliorated some of these constraints, and we are better positioned to estimate freshwater abundance and connectivity at macroscales.

In this study, we characterize the macroscale patterns of the integrated freshwater landscape that includes lakes, wetlands, and streams, across the Midwest and Northeast United States. We quantify lake, wetland, and stream abundance and density and group systems based on their surface connections using landscape position metrics. For the scope of this study, we define freshwater connectivity as the permanent surface hydrologic connections that link lakes, wetlands, and streams, and we measure connectivity as the landscape position of systems within stream networks. Our objectives were to (1) quantify macroscale patterns of surface freshwater characteristics that integrate measures of both abundance and connectivity at a subcontinental spatial extent and (2) relate these macroscale patterns to hydrogeomorphic, land-use, and climatic attributes that are hypothesized to be associated with the distribution and spatial characteristics of surface freshwaters. Within our macroscale extent (~1,800,000 km^2), we found distinct spatial patterns in the abundance and connectivity of freshwaters that can help scientists and managers develop an improved broad-scale understanding of the composition and structure of the freshwater landscape.

Methods

Overview of analytical conceptual approach

No single measure can capture the diversity and spatial complexity of surface freshwater systems, and detailed metrics requiring site-specific data inputs are impractical across broad, regional settings. Therefore, we adopted a multi-step approach to build a synthetic view of the freshwater landscape and investigate broad-scale controls on macroscale patterns. We first mapped freshwater abundance and density of lakes, wetlands, and streams separately across the study extent to determine the proportion of the landscape that is composed of surface freshwater bodies. Second, we examined spatial distributions of lake, wetland, and stream connectivity by characterizing systems based on their stream network landscape position and quantified the relative proportion area for each freshwater system type. We then combined data on abundance and connectivity for lakes, wetlands, and streams to develop an integrated view of the freshwater landscape.

Finally, we quantified associations between integrated freshwater clusters and hydrogeomorphic, land-use, and climate variables to explore underlying controls on spatial patterns at macroscales.

Study extent

The study area spanned a subcontinental extent (~1,800,000 km^2) in the temperate Midwest and Northeast regions of the United States (Fig. 1). This spatial extent presents a distinctive opportunity to capture the diversity of freshwater systems and connectivity characteristics that make up the freshwater landscape because it is rich in lakes, wetlands, and streams and spans regional settings with diverse hydrogeomorphic, land-use, past glacial regime, and climate conditions (Appendix S1: Table S1). To illustrate this, at the Hydrologic Unit 8 scale (HU8), we found a range of values for geologic composition (e.g., alluvial geology 0–89%), hydrologic attributes (e.g., mean runoff 44–762 mm/yr), land-use activities (e.g., agriculture 0–89%), and climate variables (e.g., mean precipitation 566–1376 mm/yr). In addition, the study extent has different glaciation regimes from the most recent glacial period (Wisconsin period ~2.6 million to 11,000 yr ago), with 58% of the study extent having been glaciated, 37% unglaciated, and 3% partially glaciated (i.e., area along glaciation boundary that includes glaciated and unglaciated terrain).

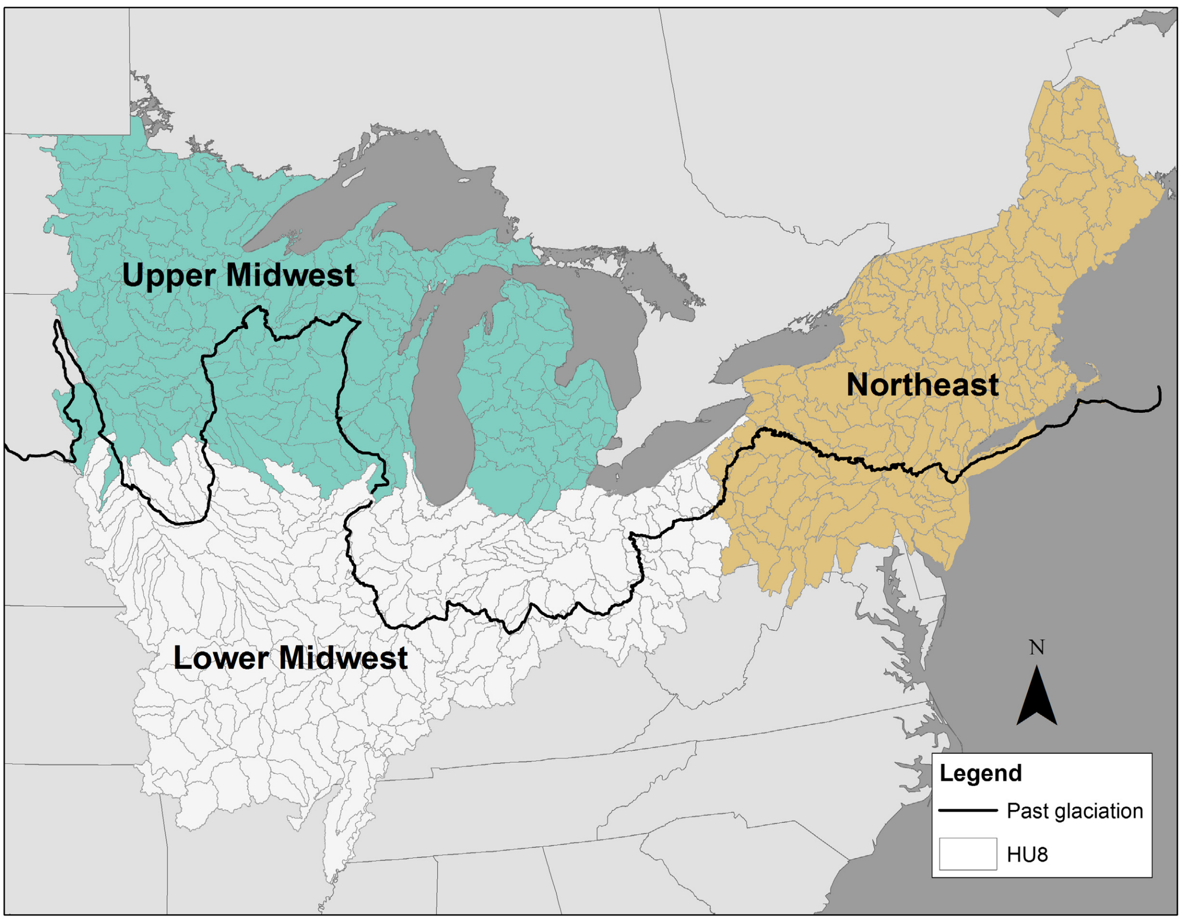

Fig. 1. Study extent regions. The study extent includes three regions in the United States: the Upper Midwest, Lower Midwest, and the Northeast and in total spans ~1,800,000 km^2 in area. Hydrologic Unit 8s (HU8, light gray polygons) are depicted on the map and have a median size of 2880 km^2. The solid black line represents the estimated boundary of the Wisconsin glacial period—north of the line is glaciated area and south of the line is unglaciated area.

Description of datasets

Our analyses used data in the LAGOS-NE_{GEO} database v. 1.03 (LAke multi-scaled GeOSpatial and temporal database), an integrated, multi-thematic geographical database spanning the Midwest and Northeast United States (Soranno et al. 2015*b*). LAGOS-NE_{GEO} includes metrics derived from national-scale geographic data on geology, topography, hydrology, climate, and land use/land cover from multiple sources (Appendix S1: Table S2). Lake and stream data originally came from the National Hydrography Dataset (NHD version 9.3; 1:24,000 resolution; United States Geological Survey [USGS], USGS Headquarters, Reston, Virginia, USA), and wetland data came from the National Wetlands Inventory (NWI version 2 Surface Waters and Wetlands Inventory; U.S. Fish and Wildlife Service, Northwest, Washington, D.C., USA). Other geospatial data in LAGOS-NE_{GEO} came from separate data sources: hydrology (USGS National Water Information System portal), topography (USGS National Elevation Dataset), geology and glaciation limits (USGS surficial materials map database), climate data as 30-yr normal averages (PRISM Climate Group), and land use/land cover (2011 National Land Cover Database, Multi-Resolution Land Characteristic Consortium). See Soranno et al. (2015*b*) for more information on the database.

Definitions of lakes, wetlands, and streams

LAGOS-NE_{GEO} included lake, wetland, and stream features from the original NHD and NWI geographic layers that captured features that were considered to be part of the surface freshwater landscape and to account for limitations in the original geographic data layers (Appendix S1: Table S3). For example, both the NHD and NWI comprehensively map freshwater feature types including artificial systems (e.g., sewage treatment ponds) and artificial flowlines (e.g., lines connecting NHD features) and these highly modified, man-made systems were not incorporated into analyses. Our freshwater system definitions are provided below.

Lake features came from NHD v. 9.3 and were defined as perennial water bodies (as denoted within the NHD) with surface area $\geq$0.04 km^2, including both natural lakes and reservoirs (i.e., impounded streams or rivers), with a total of 50,726 lakes matching our definition within the study extent. Lakes smaller than 0.04 km^2 in size were excluded from the analyses because these smaller water bodies were associated with high identification and digitization error rates compared to lakes that were 0.04 km^2 and larger in the NHD data layer (Appendix S1: Table S3; Soranno et al. 2015*b*). Omitting these small lakes (<0.04 km^2 and $\geq$0.01 km^2) in the NHD did not greatly affect lake area summaries accounting for only 0.73% of total lake area in the study extent; but, they made up over half the number of inland lake bodies (63%). However, there is evidence that lakes smaller than 0.04 km^2 have large digitization error rates (~58%) compared to lakes larger than 0.04 km^2 (~15%; Soranno et al. 2015*b*). We did not distinguish between natural lakes and reservoirs because not all systems in the NHD are properly categorized and many artificial water bodies have been mislabeled as natural lakes (McDonald et al. 2012).

Wetland features came from NWI v.2 and included NWI classified *Palustrine* systems, that is, non-tidal wetlands dominated by trees, shrubs, and persistent emergent vegetation (Cowardin et al. 1979). We removed deep-water systems (lacustrine and riverine) from the dataset and erased lake area from wetland polygons we retained. There were over 4 million wetland polygons remaining in the study extent (n = 4,411,383) following the preprocessing steps. We did not set size restrictions on wetland data because NWI reliably captures wetlands that are 0.002 km^2 in size or larger with a 98% identification accuracy rate (Wetland Mapping Standard; Federal Geographic Data Committee). We kept the NWI wetland class geographic delineations based on vegetation and substrate composition even for polygons that were adjacent to one another. While it can be argued that these adjacent polygons are hydrologically connected to one another, we chose to retain these boundaries rather than to merge them into single wetland polygons because different vegetation and substrate composition is associated with water regime and other hydrologic characteristics (Cowardin et al. 1979), which is important information we wanted to retain. However, future studies should consider how merging wetland polygons influences connectivity estimates.

Finally, stream data included perennial streams classified as *Stream/River*, *Canal/Ditch*, *Connector*, *Pipeline*, *Underground Conduit*, and *Artificial Path* feature types in the NHDFlowline data layer and

NHDArea polygons that captured wide stream and river features. We excluded *Coastlines* linear features and *Artificial Path* features that did not spatially intersect NHDArea polygons with *Stream/River* features. We did not include intermittent stream types in this analysis, although we recognize the importance of intermittent and headwater streams in aquatic ecosystem function and dynamics (Nadeau and Rains 2007). We only included perennial stream features because they likely represent more consistent hydrologic connections and because these headwater and intermittent streams are often underrepresented in the NHD in comparison with field surveys (Fritz et al. 2013). In total, stream length exceeded 1.7 million km in the study extent. Freshwater feature definitions and detailed descriptions of preprocessing steps for the data in LAGOS-NE_{GEO} are provided in Soranno et al. (2015*b*).

There were limitations in the lake, wetland, and stream geographic data that we were not able to address in our analyses (Appendix S1: Table S3). For example, both the NHD and NWI are national datasets but are produced and regularly updated at regional (often state-level) scales by local entities. As a consequence, there are differences in data resolution and digitization intensity of water bodies across the spatial extent. Freshwater features may not be evenly digitized across the study extent, which could affect freshwater abundance and connectivity estimates.

Measuring freshwater features in the landscape at macroscales

Freshwater abundance and connectivity metrics were measured at two HU spatial scales: HU12 and HU8 (Seaber et al. 1987). The HUs are hierarchically nested stream watershed spatial units that are based on USGS 1:24,000-scale topographic maps. The HU12 scale is the smallest nested spatial unit in LAGOS-NE_{GEO} and was used to examine fine-scale heterogeneity in freshwater attributes (n = 18,870; median size = 78.10 km^2). The HU8 scale is an intermediate-sized spatial unit and was used to examine regional-scale heterogeneity in freshwater attributes (n = 445; median size = 2880 km^2).

Freshwater abundance.—Freshwater abundance for lakes and wetlands was quantified as the total proportional area and density within HU12 spatial units. Streams were measured as the total stream density (length per unit area) within HU12 spatial units. We represented freshwater abundance by mapping binned metric values for lakes, wetlands, and streams separately.

Freshwater connectivity.—We characterized freshwater connectivity within the spatial extent as follows: (1) We quantified multiple freshwater connectivity metrics for lakes, streams, and wetlands separately, (2) we performed principal components analysis (PCA) on the connectivity metric values for each freshwater type to reduce collinearity, (3) we performed *k*-means cluster analysis to group spatial units with similar freshwater connectivity characteristics, and (4) we mapped the cluster assignments to visualize spatial patterns.

Freshwater connectivity metrics.—Freshwater connectivity metrics were calculated by first assigning lake, wetland, and stream geographic features into connectivity types based on their spatial arrangement with other freshwater features (Fig. 2). Lake connectivity types were based on landscape position, defined by the surface hydrologic connections of the focal lake with inflowing and outflowing streams and upstream lakes. Lakes were grouped into four hydrologic classes: Isolated, Headwater, Drainage, and Drainage-UPLK (Fig. 2a). Isolated lakes are defined as having no stream inlets or outlets. Headwater lakes have no stream inlets and at least one outlet. Drainage lakes have inlets and outlets and no upstream lakes (≥0.10 km^2 in size). Drainage-UPLK lakes have inlets and outlets and at least one upstream lake (≥0.10 km^2 in size).

Wetland connectivity types also were defined in relation to connections with surface stream networks; however, streamflow directionality was not incorporated into the definition due to computation limitations of processing millions of wetland polygons. Wetlands were grouped into three classes: Upland-Embedded, Headwater, and Drainage (Fig. 2b). We considered wetlands to be stream-connected if located at any point within a 30-m buffer surrounding the stream segment center line or double-banked lines for large rivers (Soranno et al. 2015*b*); this buffer approach was necessary to accommodate spatial data resolution limitations and misalignment between different data layers (NHD and NWI). We called wetlands with no permanent surface stream connections Upland-Embedded. These systems are thought to

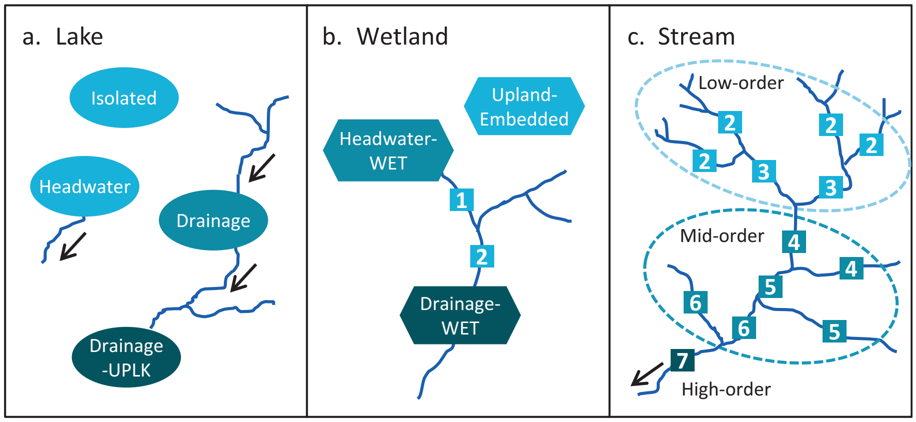

Fig. 2. Diagram of freshwater connectivity metrics for (a) lakes, (b) wetlands, and (c) streams. (a) Lakes (oval) were assigned connectivity types based on spatial relationships with streams and upstream lakes. Lake connectivity types include Isolated, no stream inlets or outlets; Headwater, only stream outlets; Drainage, stream inlets and outlets; and Drainage-UPLK, stream inlets and outlets and at least one upstream lake ≥0.10 km^2 in size. (b) Wetlands (hexagon) were assigned connectivity types based on spatial relationships with streams. Wetland connectivity types include Upland-Embedded, no intersecting stream segments; Headwater (Headwater-WET), intersected by one first-order stream segment; and Drainage (Drainage-WET), intersected by a stream segment > first-order or multiple stream segments. (c) Stream segments were assigned Strahler stream order (squares) based on location in the stream network and grouped into stream connectivity classes: Low-order (first–third), Mid-order (fourth–sixth), and High-order (>sixth).

have low connectivity with no persistent surface hydrologic connections but potentially intermittent and/or subsurface hydrologic connections in the landscape (Tiner 2003, Mushet et al. 2015). Because flow directionality of streams intersecting wetlands was not examined and inlets and outlets could not be identified, we defined Headwater wetlands to be systems that are intersected by a Strahler first-order stream segment. And finally Drainage wetlands were defined as systems intersected by either a single second-order or higher stream segment or by multiple stream segments.

Stream features in our study extent were classified by Strahler stream order, a measure of the position of the stream reach in relation to the stream network with order increasing as one moves from the headwater streams to the terminal point (Strahler 1957). We considered stream order to capture differences in stream connectivity characteristics based on the assumption that low-order streams are in headwater watersheds that potentially could have more intermittent flows as compared to high-order streams that may have more persistent connections. Stream order classes were grouped into three categories: Low-order, Mid-order, and High-order (Fig. 2c). Low-order streams include first- to third-order stream reaches, Mid-order streams include fourth- to sixth-order reaches, and High-order streams include greater than sixth-order reaches.

We calculated the relative proportion of each connectivity type to the total area or length of the respective freshwater system in a spatial unit (e.g., Isolated lake area divided by total lake area within the HU12). Relative proportion metrics are better suited to our connectivity analyses as compared to total areal proportion metrics (e.g., Isolated lake area divided by HU12 area) because freshwater features cover only a small fraction of area across our macroscale extent, resulting in many observations having zero or very low values. More information on geoprocessing steps and the GIS toolbox developed to calculate the connectivity metrics is described in Soranno et al. (2015*b*), and the toolbox is available at https://github.com/soranno/LAGOS_GIS_Toolbox.

Analysis

Principal components and k*-means cluster analyses for freshwater connectivity.*—Prior to analyses, we transformed the relative proportion (logit) and stream density (natural log) metric values to meet assumptions of normality and homoscedasticity. We performed PCAs on the transformed connectivity metrics that were quantified at the HU12 scale for lakes, wetlands, and streams separately to reduce codependence among values. Principal components analyses were performed using JMP 10 software (SAS Institute, Cary, North Carolina, USA). Due to co-variation among the variables, the top two axes explained between 70% and 98% of the variation in the different PCAs performed; we only used the scores of these two first axes for the subsequent cluster analysis. The first two axes of the PCA are shown in Appendix S1: Fig. S1.

K-means analyses were performed to group HU12s that had similar freshwater connectivity characteristics, based on PCA scores, using JMP10. Prior to analyses, we used the "Cascade *k*-means" approach (vegan R package; Oksanen et al. 2017) to identify the number of clusters that optimizes the within- vs. among-cluster residual mean square error. There were some spatial units that were not assigned cluster groups because there were no freshwater systems present (e.g., HU12s without lakes present $n = 7922$ and comprised ~39% of study extent area) or because stream segments were not assigned Strahler stream order class due to digitization errors in streamflow directionality in the original NHD data layer ($n = 123$ HU12). These sites were assigned their own cluster, named "zero lake" and "no stream order." Finally, we mapped the cluster groups across the study extent to visualize the patterns of freshwater surface connectivity at macroscales.

Patterns of abundance and connectivity integrated across lakes, wetlands, and streams.—To visualize regional patterns of abundance and connectivity integrated across lakes, wetlands, and streams, we performed a similar set of analyses as the HU12 freshwater connectivity analysis (i.e., PCA, cluster analysis, and mapping the cluster groups) at the HU8 scale. We included abundance and connectivity metrics across the three freshwater systems in the PCA because it is expected that these freshwater features may co-vary with one another and be influenced by similar geophysical and climatic drivers. We used the HU8 scale because lakes were present in only 58% of the HU12s; and 60% of HU12s with lakes only had 1–3 lakes, which could result in lakes (or the lack of lakes) playing a disproportionate role in determining broad-scale freshwater patterns. Performing the integrated analyses at the HU8 scale allowed all spatial units to include lakes, providing a more balanced weight to lakes, wetlands, and streams. Eight HU8 spatial units were not assigned cluster groups because they contained streams that were not assigned Strahler stream order class. In total, 447 HU8 spatial units (out of 455) were retained to create integrated freshwater clusters across the study extent.

We examined the geospatial composition of the resulting integrated freshwater clusters by creating boxplots of freshwater abundance and connectivity metrics and other landscape features grouped by cluster. We tested for differences in geospatial composition among cluster groups using non-parametric Kruskal–Wallis tests and using multiple comparison tests to examine differences between individual clusters ($\alpha < 0.05$).

Hydrogeomorphic, climate, and land-use variables associated with freshwater abundance and connectivity.—We performed random forest analyses using randomForest package in R (Liaw and Wiener 2002) to examine associations between the integrated freshwater cluster groups (at HU8 scale) and underlying landscape and climatic characteristics (Appendix S1: Table S1). Random forest is a machine-learning technique based on classification regression trees and combines multiple classification trees to improve classification accuracy (Cutler et al. 2007). The algorithm uses bootstrap samples of the original observations and randomized subsets of predictor variables to build individual trees. Model accuracy and variable importance were estimated from the hold-out observations (about one-third of the observations in the training set; Breiman 2001). We used the VSURF package to inform variable selection in the final random forest model (Genuer et al. 2016). The package provides an automated method to select variables based on importance scores and to minimize redundancy among predictor variables and thus allowed us to identify the geographic metrics that were most closely associated with freshwater abundance and connectivity clusters. The number of bootstrap samples was fixed to the smallest cluster group size to improve prediction accuracy among unbalanced cluster groups and to reduce bias of minimizing error for large cluster groups.

Freshwater clusters, PCA scores, and freshwater metric and landscape composition data at the HU12 and HU8 scales are provided in the Environmental Data Initiative (https://environmentaldatainitiative.org/; Fergus et al. 2017*a*, *b*).

Results

Freshwater abundance

Lake, wetland, and stream abundance measured as areal proportions or stream density within the HU12 exhibited contrasting macroscale patterns within the study extent (Table 1, Fig. 3a–c). Lakes (≥0.04 km^2 in size) were only present in over half of the study extent (61%) in contrast to wetlands and streams that were found throughout the study extent (~99%). In total, lake area made up 1.9% of the study extent area (with 2.2% lake area in glaciated areas and 0.5% in non-glaciated areas), and lake density was 0.03 km^{-2}. In contrast to lakes, wetlands made up 7.8% of the total study extent area and were found at a higher density of 2.8 km^{-2}. Total stream length was 1,794,044 km with an overall density of 1.13 km/km^2 in the study extent. In general, lake and wetland abundance exhibited opposite spatial patterns along a north-to-south gradient compared to stream density. Lake and wetland abundance was higher in

Table 1. Summary statistics of freshwater metrics.

Freshwater metric type	Metric	Median (SD)	Range
HU8 scale			
Lake abundance	Lake proportion	0.01 (0.04)	0–0.41
Lake size	Lake size (km^2)	0.48 (803)	0–1,173,449
Lake–stream connectivity	Isolated	0.08 (0.18)	0–1.00
Relative proportion	Headwater	0.05 (0.11)	0–0.95
	Drainage	0.30 (0.26)	0–1.00
	Drainage-UPLK	0.41 (0.31)	0–1.00
Wetland abundance	Wetland proportion	0.05 (0.11)	0–0.95
Wetland size	Wetland size (km^2)	0.02 (0.04)	0–44.35
Wetland–stream connectivity	Upland-Embedded	0.27 (0.14)	0.02–0.68
Relative proportion	Headwater-WET	0.11 (0.05)	0–0.27
	Drainage-WET	0.59 (0.17)	0.22–0.97
Stream abundance	Stream density (m/km^2)	0.11 (0.05)	1.95–45.12
Stream connectivity	Low-order	0.87 (0.04)	0.71–1.00
Relative proportion	Mid-order	0.12 (0.04)	0–0.29
	High-order	0.001 (0.02)	0–0.13
HU12 scale			
Lake abundance	Lake proportion	0.001 (0.05)	0–0.99
Lake size	Lake size (km^2)	0.18 (5665)	0–82,139
Lake–stream connectivity	Isolated	0 (0.35)	0–1.00
Relative proportion	Headwater	0 (0.26)	0–1.00
	Drainage	0.29 (0.40)	0–1.00
	Drainage-UPLK	0 (0.35)	0–1.00
Wetland abundance	Wetland proportion	0.03 (0.11)	0–0.99
Wetland size	Wetland size (km^2)	0.02 (0.11)	0–425
Wetland–stream connectivity	Upland-Embedded	0.26 (0.23)	0–1.00
Relative proportion	Headwater-WET	0.09 (0.10)	0–1.00
	Drainage-WET	0.59 (0.26)	0–1.00
Stream abundance	Stream density (m/km^2)	0.11 (0.05)	0–70.27
Stream connectivity	Low-order	0.90 (0.13)	0–1.00
Relative proportion	Mid-order	0.09 (0.12)	0–1.00
	High-order	0 (0.05)	0–0.99

Notes: SD, standard deviation. Lake, wetland, and stream abundance and connectivity metrics were quantified at two Hydrologic Unit (HU) spatial scales: HU8 (n = 447) and HU12 (n = 18,876). Lake and wetland connectivity metrics were calculated as relative proportions—the area of each connectivity type divided by the total lake or wetland area within a spatial unit. Stream connectivity was calculated as the relative density of stream connectivity type out of the total stream length within the spatial unit.

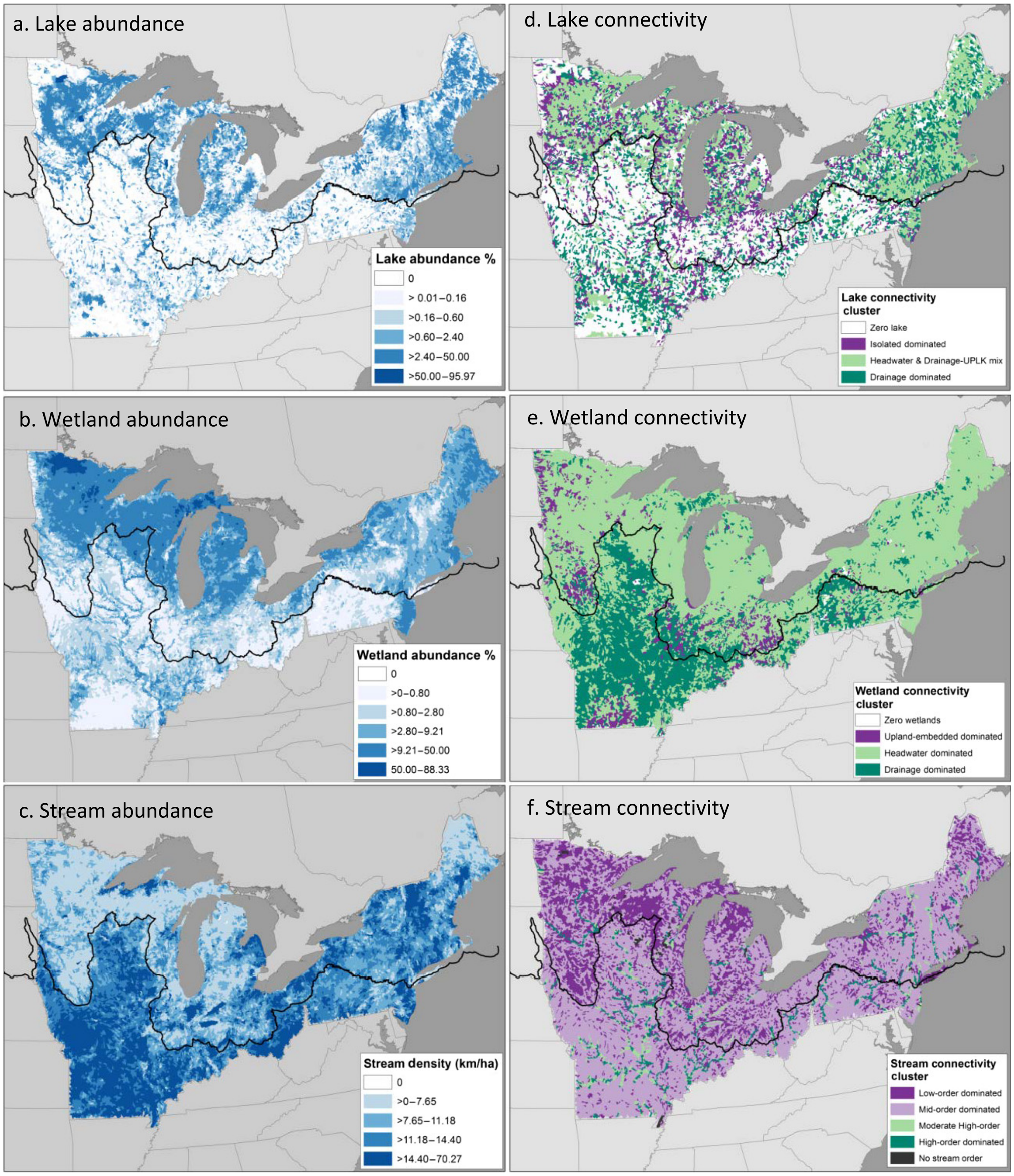

Fig. 3. Freshwater abundance and connectivity maps by system type**.** Freshwater abundance is quantified as the total proportion area or stream density within the Hydrologic Unit (HU) 12 spatial unit for lakes (a), wetlands (b), and streams (c). Abundance values are binned as quantiles. Freshwater connectivity for lakes (d), wetlands (e), and streams (f) is represented by connectivity cluster groups determined by *k*-means cluster group using principal components analysis (PCA) scores from lake, wetland, and stream connectivity metrics at the HU12 spatial scale. Dominated is in reference to where spatial units plotted on the PCA axes using relative proportion connectivity metric values. The solid black line represents the estimated boundary of the Wisconsin glacial period—north of the line is glaciated area and south of the line is unglaciated area.

the northern area of the study extent compared to the southern areas, whereas stream density was higher in the southern area of the study extent compared to northern areas. However, there were exceptions to this general trend with glaciation extent. Lakes were not present and wetlands were in low abundance in parts of the Upper Midwest (Fig. 3a, b), and stream density in the Northeast did not show as strong a north–south trend (Fig. 3c).

Freshwater connectivity

Lake, wetland, and stream connectivity clusters assigned at the HU12 scale were derived separately from PCA scores. For lakes, two principal component axes explained ~70% of variation in lake connectivity metrics and resulted in three lake connectivity clusters. For wetlands, ~96% of variation in wetland connectivity metrics was explained by two PC axes, and PCA scores resulted in three wetland connectivity clusters. For streams, ~99% of variation in stream order density metrics was explained by two PC axes, and four stream connectivity clusters were retained (Appendix S1: Fig. S1).

Spatial patterns in freshwater connectivity clusters (Fig. 3d–f) differed from corresponding freshwater abundance patterns. Across the study extent, lake connectivity clusters exhibited non-contiguous spatial patterns (Fig. 3d). For example, in the Midwest where lake abundance was high, connectivity clusters included Headwater and Drainage-UPLK lake clusters and Isolated lake clusters. In contrast, wetland connectivity clusters were contiguous and similar to wetland abundance patterns (Fig. 3e). Clusters characterized by Headwater wetlands were the most prominent cluster group (66% of HU12s) and tended to be located in the northern extent of the study area where wetland abundance was high. Areas where wetland abundance was low tended to be dominated by Drainage or Upland-Embedded wetlands. For streams, Low-order streams were prominent in the Upper Midwest HU12s and tended to be in areas where stream density was high.

The freshwater landscape: integrating lake, wetland, and stream abundance and connectivity

We combined freshwater abundance and connectivity measures for lakes, wetlands, and streams that were quantified at the HU8 spatial unit (Table 1) to evaluate the integrated freshwater landscape. Due to the higher number of variables included and the weaker co-variation when different variable types are included, the first two axes of the PCA performed on all the abundance and connectivity metrics captured ~43% of variation in the freshwater metrics (Appendix S1: Fig. S2), lower than the variation captured in the individual HU12 connectivity analyses on lake, wetland, and stream systems (Appendix S1: Fig. S1). Nevertheless, the first two axes of the PCA performed on 15 different input variables captured nearly half of the variability, showing that there are consistent patterns across connectivity, abundance, and system types that can be summed into synthetic PCA axes. Thus, based on multivariate associations among these metrics (PCA scores), we identified five integrated freshwater cluster groups with distinct freshwater abundance and connectivity attributes (Table 2; Appendix S1: Fig. S2). There were significant

Table 2. Composition of integrated freshwater abundance and connectivity cluster groups at the HU8 scale.

	Abundance			Connectivity									
				Lake				Wetland			Stream order		
Cluster assignment	Lake	Wet	Stream	I	HW	DR	UPLK	U	HW	DR	Low	Mid	High
A	–	–	+			+				+	+		+
B	–	–	+			+				+			+
C	+	+	–	+	+			+	+			+	
D	+	+	–				+		+		+		
E	+	+	–				+	+			+		

Notes: Integrated freshwater cluster groups (A–E) are composed of different freshwater abundance and connectivity characteristics. The proportion of lake and wetland area and stream (density) in clusters ranges from high (+) and low (–) abundance. Freshwater connectivity types that are dominant in clusters are represented with (+). Freshwater connectivity types include lakes and wetlands: I, Isolated; U, Upland-Embedded; HW, Headwater; DR, Drainage; UPLK, Drainage-UPLK; and streams Low = first- to third-order stream segments; Mid = fourth- to sixth-order stream segments; and High = >sixth-order stream segments.

differences (α < 0.05) in freshwater abundance and connectivity metrics among the clusters based on Kruskal–Wallis tests (Appendix S1: Table S4). However, not all cluster groups were significantly different from one another for a given metric (Appendix S1: Table S5). For example, Clusters C and D were not significantly different in lake, wetland, and stream abundance composition from one another, but they were significantly different from the lake, wetland, and stream abundance composition in Clusters A and B.

These integrated freshwater clusters exhibited macroscale spatial patterns that were largely distinguished by glaciation extent boundaries (Fig. 4). In general, we found that Clusters C, D, and E mostly located north of the glaciation boundary had significantly greater lake and wetland abundance compared to the other clusters and were composed of a variety of both isolated, upland-embedded, and stream-connected lake and wetland connectivity types, whereas Clusters A and B mostly located south of the glaciation boundary had significantly greater stream density compared to glaciated areas and were dominated by Drainage connectivity types for both lakes and wetlands (Figs. 4, 5; Appendix S1: Table S5).

The most common integrated freshwater cluster group in the study extent was Cluster C (149

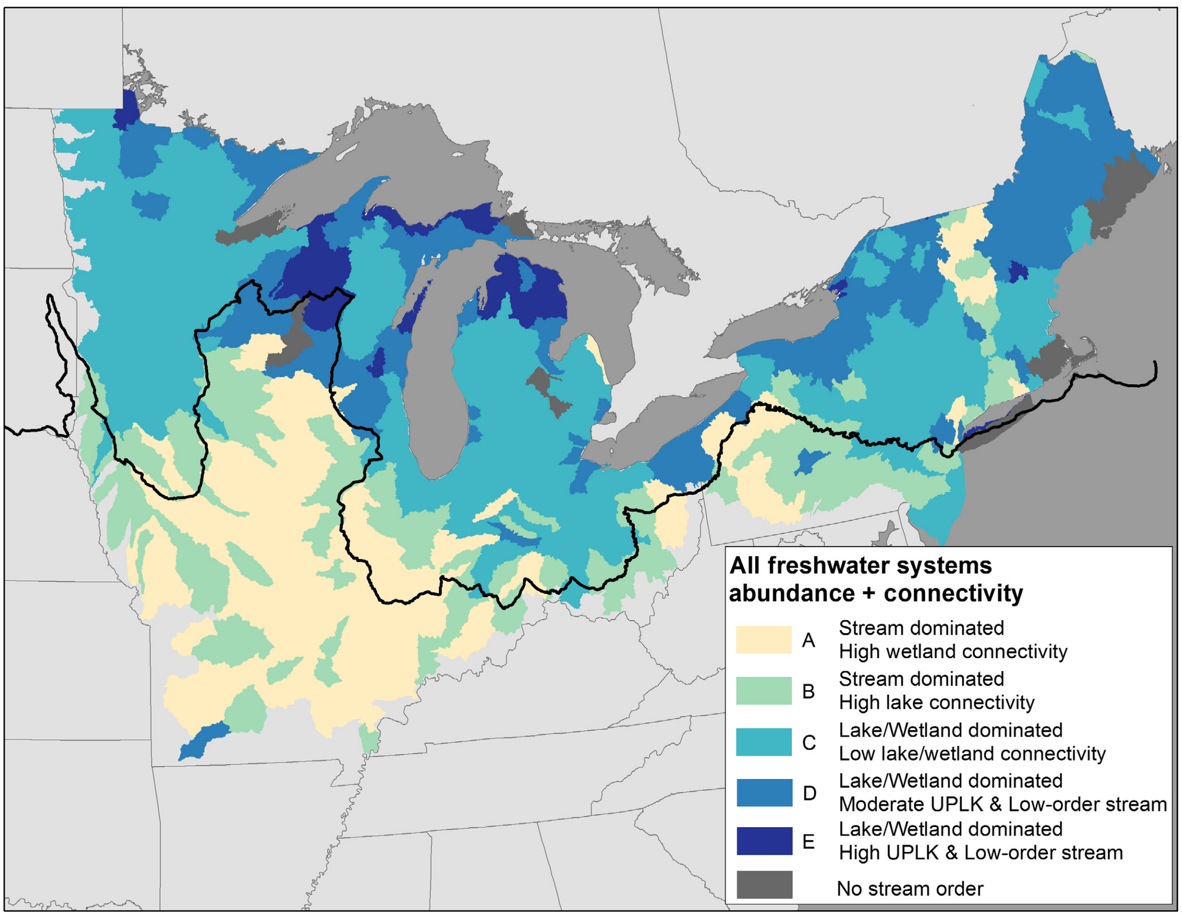

Fig. 4. Integrated freshwater abundance and connectivity map. Integrated freshwater abundance and connectivity clusters were assigned using *k*-means cluster analysis based on principal components analysis scores from lake, wetland, and stream abundance and connectivity metrics quantified at the Hydrologic Unit (HU) 8 scale (HU8 = 447). Spatial units not assigned a cluster group were missing stream order data for a portion of the area. Interpretation of cluster groups is provided in Table 1. The solid black line represents the estimated boundary of the Wisconsin glacial period—north of the line is glaciated area and south of the line is unglaciated area.

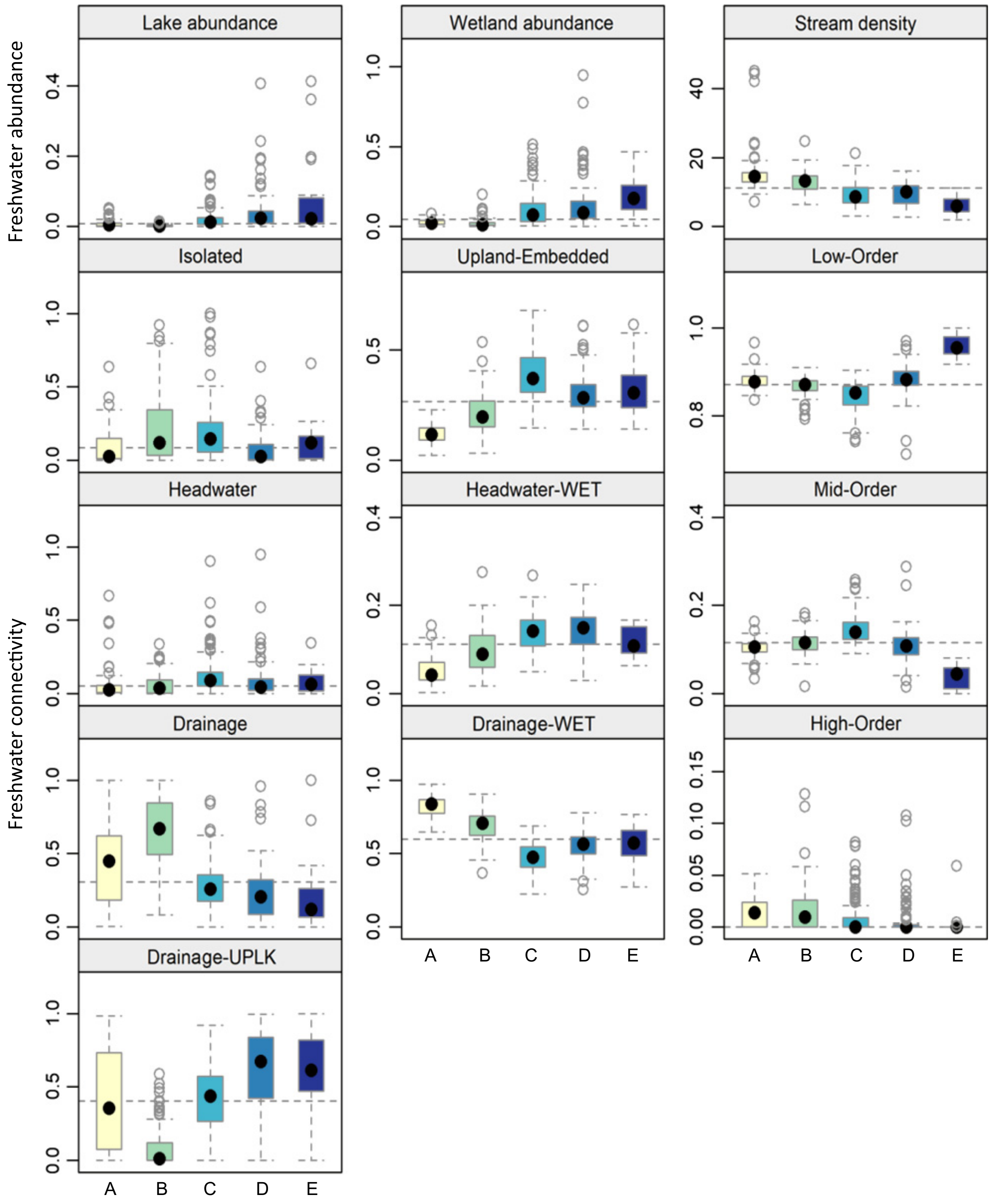

Fig. 5. Boxplots of freshwater abundance and connectivity metrics by the integrated freshwater cluster groups. Boxplots represent the distributions of lake and wetland abundance (proportion area in Hydrologic Unit [HU] 8 unit), stream density (m/km^2), and freshwater connectivity measures (relative proportion of freshwater connectivity type among respective freshwater system total area or density) among the integrated freshwater HU8 clusters (A–E). The dashed gray line represents the median metric value. Integrated freshwater clusters captured significant differences in freshwater abundance and connectivity metrics based on Kruskal–Wallis tests ($\alpha < 0.05$).

out of a total of 447 HU8s). This cluster group was distributed across the Upper Midwest and Northeast (Fig. 4) and was characterized as having significantly greater relative proportions of Isolated, Upland-Embedded, and Headwater lakes and wetlands and lower stream density that was dominated by Mid-order streams compared to the other cluster groups—except that Clusters C and B were not significantly different for Isolated lake metrics (Fig. 5; Appendix S1: Tables S4, S5). The least common integrated freshwater cluster group, Cluster E (23 out of 447 HU8s), was located primarily in the Upper Midwest. This cluster group was characterized as having significantly higher relative proportions of Low-order stream length compared to other cluster groups (Appendix S1: Table S5). However, Cluster E included HU8s with large variation in freshwater composition, and this cluster group did not significantly differ in lake and wetland connectivity metrics from other clusters (Appendix S1: Table S5).

The integrated freshwater landscape exhibited distinct, non-random spatial patterns at broad extents which provide a framework for investigating macroscale relationships that influence freshwater distributions and their connections with the surrounding landscape.

Hydrogeomorphic, climate, and land-use variables associated with broad-scale freshwater landscape attributes

Freshwater abundance patterns aligned with past glacial activity, but freshwater connectivity patterns did not completely follow this spatial gradient, suggesting that there are additional underlying drivers shaping the freshwater landscape. Of the 21 predictor variables tested in random forest models analyzing integrated freshwater abundance and connectivity clusters (HU8 scale), glaciation regime was the top predictor. In fact, there were significant differences in lake, wetland, and stream abundance metrics across glaciation regimes (Kruskal–Wallis tests: lake abundance, $H = 93.69$, df = 2, $P < 0.0001$; wetland abundance, $H = 133.91$, df = 2, $P < 0.0001$; stream density, $H = 138.79$, df = 2, $P < 0.0001$). But Isolated lake metrics were not significantly different among glaciation regimes ($H = 4.28$, df = 2, $P = 0.12$), and Low-order stream metrics were only marginally significantly different among glaciation regimes ($H = 8.50$, df = 2, $P = 0.01$). The associations suggest that past geological activity (i.e., glacial processes) is likely a key driver that affects the presence and abundance of freshwater systems, while specific hydrologic, geologic, and human land-use activity may influence freshwater connectivity within regions in different ways.

The top performing random forest model accurately predicted 57% of integrated freshwater cluster assignment (HU8 scale) based on out-of-bag samples, meaning that the predictor variables did better than random at predicting the overall cluster membership. The most important variables associated with the integrated clusters were glaciation regime, hydrology (mean runoff and baseflow), geology (glacial fluvial outwash and alluvial deposits), mean precipitation, and human land use (pasture and agriculture; Appendix S1: Fig. S3). These variables were significantly different among the five integrated cluster groups based on Kruskal–Wallis tests (Appendix S1: Table S6). Prediction accuracy varied among individual integrated freshwater clusters with Cluster A having the highest classification accuracy (65%) and Cluster B having the lowest (45%; Appendix S1: Table S7).

The integrated freshwater clusters differed in hydrogeomorphic features, climate, and land use (Fig. 6). In general, areas that were lake and wetland rich (Clusters C, D, and E) tended to have high runoff, baseflow, and glacial fluvial outwash geology. In contrast, areas that were rich in streams (Clusters A and B) tended to have high precipitation, topographic slope, alluvial geology, and pasture and agricultural land-use activities. The location of integrated freshwater clusters tended to vary with glaciation regime with Clusters A and B observed in higher frequency in unglaciated areas and Clusters C, D, and E observed in higher frequency in glaciated areas (Appendix S1: Fig. S4). Although glaciation regime co-varied with other geospatial predictors such as mean baseflow, mean runoff, percent agriculture, and mean slope based on Kruskal–Wallis tests (Appendix S1: Fig. S5), redundancy analyses in the variable selection step (VSURF package) retained these variables, indicating that they contained independent information from glaciation regime.

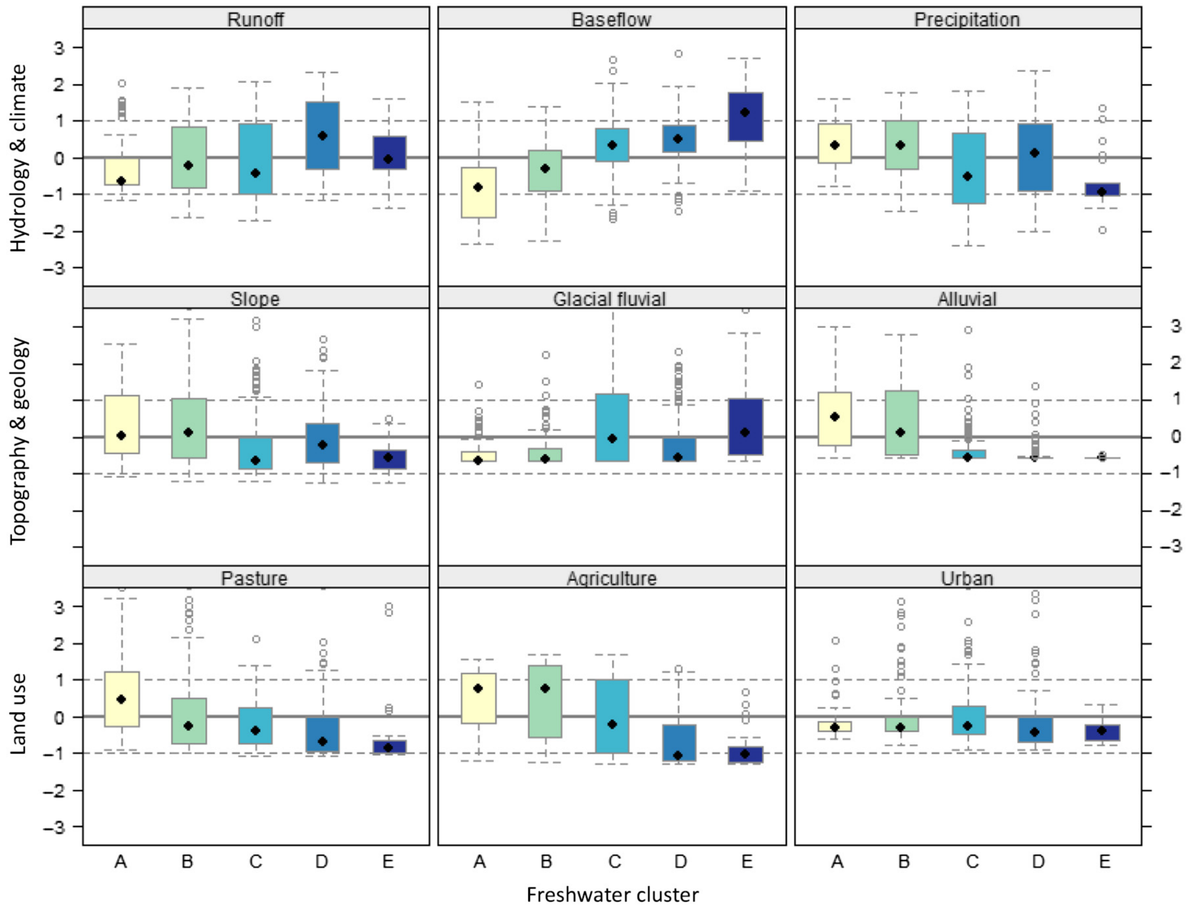

Fig. 6. Boxplots of the geophysical predictors of the integrated freshwater clusters. Boxplots represent the standardized distributions of geophysical variables among the integrated freshwater Hydrologic Unit 8 clusters (A–E). Geophysical variables were the top predictors of the integrated freshwater clusters in random forest analyses and are grouped as hydrology and climate, topography and geology, and land use. Values were standardized by subtracting the mean and dividing by the standard deviation. The solid gray line represents the mean, and the dashed lines represent ±1 standard deviation above or below the mean. There were overall significant differences in hydrology, climate, topography, geology and land-use values among the clusters based on Kruskal–Wallis tests ($\alpha < 0.05$).

Discussion

We present a synthetic view of the integrated freshwater landscape. Our analyses of lakes, wetlands, and streams illustrate patterns and diversity of freshwater systems and surface connectivity at macroscales. We found four important results. (1) Abundance of lakes and wetlands exhibit spatial patterns that are opposite from stream density and that mostly follow glaciation extent boundaries. Lake and wetland abundance is higher in glaciated areas compared to unglaciated areas, whereas stream density is lower in glaciated areas compared to unglaciated areas. (2) We found distinct, broad-scale patterns among lake, wetland, and stream connectivity that reveal a layer of complexity that abundance measures alone did not capture, suggesting that freshwater abundance and connectivity may be influenced by different underlying processes. (3) There were spatially contiguous patterns in abundance and connectivity measures across all three freshwater system types (i.e., lakes, wetlands, and streams). (4) These patterns in abundance and connectivity were associated with underlying hydrogeomorphic, climate, and land-use variables. We have illustrated a robust approach to quantitatively measure freshwater abundance and connectivity that can be incorporated into models at macroscales. These results can inform both fundamental and applied scientific questions to better understand and manage freshwater systems.

The freshwater landscape: lake, wetland, and stream abundance and connectivity

The freshwater abundance and connectivity analyses presented here provide a broad-scale, integrated picture of the freshwater landscape. The patterns and distributions of the integrated freshwater landscape (Figs. 3, 4) can inform empirical and applied sciences alike. While it is common for freshwater systems to be studied in isolation within disciplinary boundaries, it is becoming widely recognized that systems need to be studied together and that their hydrologic connections are important to ecosystem attributes and processes. Our analyses further our understanding of the freshwater landscape by incorporating freshwater surface connections into a macroscale framework. By ignoring freshwater connectivity, we lose information on structural attributes of freshwater systems that can be critical to key ecological processes. These connections are important to advance understanding in biogeochemistry (Strayer et al. 2003, Fisher et al. 2004, Racchetti et al. 2010), aquatic population and community dynamics (Crump et al. 2007), and to more accurately incorporate freshwater systems in regional and global carbon and nutrient budgets (Cardille et al. 2007, Butman et al. 2016).

Freshwater abundance estimates at regional and global extents have been conducted on lakes, wetlands, and streams separately through inventories (Lehner and Döll 2004, McDonald et al. 2012, Verpoorter et al. 2014) or extrapolation methods using size-distribution scaling laws (Meybeck 1995, Downing et al. 2006, 2012). Our results were in agreement with the spatial patterns and range of values reported in the freshwater abundance literature. Mean lake areal percentages estimated at the HU12 spatial unit were 2.2% in glaciated areas and 0.5% in unglaciated areas in comparison with estimates for the entire United States described by Meybeck (1995; 2.8% in glaciated areas and 0.09% in unglaciated areas). Mean wetland areal percent estimated at the HU12 was 7.8%, slightly larger than total wetland area (including freshwater, estuarine, and marine systems) estimate for the conterminous United States of 5.5% (Dahl 2006); but our study extent did not include arid and mountainous regions, which would likely lower wetland abundance estimates. We also found that freshwater abundance patterns generally followed the Wisconsin glaciation boundary but with inverse patterns for lakes and wetlands compared to streams. Other studies support that lake abundance (for lakes ≥1 km^2) is greater in glaciated areas compared to unglaciated areas (Meybeck 1995, Lehner and Döll 2004), but distributions in relation to glaciation regime have not been explicitly examined for wetlands and streams.

In addition to abundance estimates, we observed general freshwater connectivity patterns that are supported by previous studies conducted at other scales. Isolated and Upland-Embedded lakes and wetlands tended to be smaller in size (median size 0.001 and 0.0001 km^2, respectively) compared to Drainage lakes and wetlands (median size 0.002 and 0.0004 km^2, respectively) in the spatial extent. In fact, there is an increasing probability of a wetland being geographically isolated from surface connections with decreasing patch size (Cohen et al. 2016). The association between freshwater connectivity characteristics and water body size suggests that freshwater connectivity types may follow size-frequency scaling distributions, similar to lake abundance trends (Downing et al. 2006), to some degree, which could aid in extrapolation methods to estimate connectivity at macroscales.

In addition, we found that Low-order (first–third) streams dominated total stream length (88%) in the study extent, and the proportion of stream length declined with increasing stream order with Mid-order streams making up 11% and High-order streams making 1% of the total stream length. These trends support global and regional stream density estimates that indicate that lower-order streams have the greatest number of stream segments and make up the most total stream length (Rodriguez-Iturbe et al. 1994, Butman and Raymond 2011, Downing et al. 2012). Understanding stream network patterns are critical to upscale freshwater processes to broad extents because stream network structure can influence biogeochemical processes (Fisher et al. 2004) such as CO_2 dynamics (Humborg et al. 2010).

Other studies have characterized hydrologic settings at broad spatial extents as we have done here, but they did not explicitly use freshwater surface features, nor connectivity characteristics as data inputs. Hydrologic regions delineated within the United States were based on well-understood topography, geology, and climate characteristics and were designed to identify similar hydrologic settings for surface waters and groundwaters in the landscape (Winter 2001, Wolock et al. 2004). The objective of these hydrologic regions was to identify combinations of topographic, geologic, and climatic conditions that interact to affect hydrologic systems in the landscape; the predictors best describing connectivity reflect those used to develop hydrologic regions. Our study differed from the deductive approach used by Wolock et al. (2004) in that we used an inductive approach based on freshwater metrics rather than land- and climate-based predictors to examine spatial patterns in connectivity. Our approach complements these other hydrologic studies in that we provide a framework to study the macroscale patterns of freshwater systems and their connectivity by directly using data on lakes, wetlands, and streams as inputs in the analyses. With this framework, we can visualize where freshwater systems and connectivity attributes are located on the landscape and how they may interact with other geospatial features to affect ecological attributes and processes.

Geospatial associations with the freshwater landscape

Freshwater composition and connectivity characteristics in the landscape are shaped by a suite of drivers that influence landform (past geologic activity, topographic relief), water source (climate and hydrology), permeability of substrate to hold water (surficial geology), and by human modifications (e.g., land-use activities) that can affect all of the above. Our analysis identified hydrogeomorphic, climate, and land-use variables that may be related to these different drivers. It is expected that several of these landscape features are spatially correlated with one another, which makes it challenging to identify individual drivers of freshwater abundance and connectivity patterns. However, the variables have unique information from one another, support hypothesized relationships from the literature, and provide insight into the broad-scale variables that affect the freshwater landscape. Glaciation regime was one of the top predictors of the integrated freshwater clusters. The most recent glacial period (Wisconsin stage) scoured the landscape to create depressions and sedimentary deposits, and glaciated areas are associated with increased lake and wetland abundance (Meybeck 1995, Winter 2000, Lehner and Döll 2004). With over half of our study extent being glaciated (59%), it was expected that glaciation regime would be significantly related to patterns in the freshwater landscape and associated with increased lake and wetland abundance.

However, glaciation was not an exclusive driver of the freshwater landscape. Within glaciated areas, there can be a great deal of variability in the abundance and types of freshwater bodies present due to differences in rock type and relief (Meybeck 1995, Winter 2000). In our study extent, surficial geology was important to predict freshwater clusters and indicated past geologic activity and current hydrologic conditions. High glacial fluvial outwash deposits were associated with lake- and wetland-rich areas. This substrate is composed of sediment, sand, and gravel deposits from past glacial activity and is associated with modern groundwater aquifers. Stream-rich areas had high alluvial deposits (clay, silt, sand, and gravel) associated with past geomorphic processes, and the permeability of these geologic materials can influence surface and subsurface water exchange (Winter 2000, Fisher et al. 2004) and shape the freshwater landscape. We also found that hydrologic variables were associated with the integrated freshwater clusters—with lake- and wetland-rich areas tending to have higher runoff and baseflow compared to stream-rich areas. These variables can be indicators of surface and groundwater sources to freshwater systems. Climate variables were also associated with the freshwater landscape: Mean precipitation was significantly higher in areas with high stream densities (Clusters A and B) compared to areas with low stream densities (Clusters C and E; Fig. 6; Appendix S1: Table S6). This trend was also observed at a global scale where a positive correlation between stream abundance and precipitation was observed (Raymond et al. 2013).

Finally, we found an association between the human land use in the HU8 and the integrated freshwater cluster groups. Stream-rich clusters were associated with higher proportions of mean agriculture compared to lake- and wetland-rich areas (Fig. 6; Appendix S1: Table S6). Agricultural activity can directly modify the freshwater landscape by diverting or extracting water, creating impoundments, and draining wetlands (Smith et al. 2002, Wright and Wimberly 2013) and may preferentially remove geographically isolated (e.g., Upland-Embedded) systems (Cohen et al. 2016, Rains et al. 2016). However, agriculture may be correlated with other variables such as topography and soil composition that may influence freshwater abundance and connectivity attributes.

The hydrogeomorphic, climate, and land-use variables identified in the analyses may be indicators of the diverse drivers that shape freshwater composition and connectivity characteristics across the landscape. At macroscales, past geological activity (i.e., glacial and fluvial processes) is a key broad-scale driver that affects the presence of freshwater systems with hydrologic, geologic, and human land-use activity influencing smaller-scale variability in freshwater characteristics including connectivity across regions. It should be noted that our objectives were to identify potential geographic variables that may be associated with freshwater abundance and connectivity characteristics. The associations do not imply causative relationships but rather highlight potential variables to examine in greater depth, at finer spatial scales in relation to the freshwater landscape, particularly in the face of global change in the coming decades and centuries.

Future needs for freshwater metric developments

Our freshwater connectivity metrics are simplified representations of the spatial configuration of freshwater systems in the landscape and thus do not capture all aspects of freshwater connectivity. In particular, data are lacking on groundwater, reservoirs, and dams at broad geographic extents. Also, the metrics are only as good as the resolution and accuracy of the original data layers, and our metrics do not capture temporal changes in freshwater connectivity. Surface hydrologic connectivity is a spatial characteristic, but the magnitude and presence of these connections are dynamic through time based on seasonal changes in climate and hydrology. This is especially true for upland-embedded wetlands and intermittent and ephemeral streams. These systems lack persistent surface water connections producing conditions that support unique biogeochemical and biological functions (Larned et al. 2010, Datry et al. 2014, Cohen et al. 2016, Rains et al. 2016). But isolated and ephemeral systems are being modified and lost at high rates due to land-use activities and climate change, and thus, there is a clear need to assess what is currently present (Larned et al. 2010, Rains et al. 2016). Our metrics are based on snapshots of surface water features represented in the NHD and NWI that are based on remotely sensed imagery, and thus lack information on connectivity changes over time. Temporal changes in connectivity could be assessed by incorporating land-cover data measured at other time periods (Pekel et al. 2016) or perhaps modeling hydrologic flow dynamics using topographic and other geographic data layers, but such analysis is beyond the scope of the current study.

Our freshwater connectivity metrics represent the complex surface hydrologic network that links lakes, wetlands, and streams and moderates the flow of water, materials, nutrients, and organisms across the landscape. These metrics are relatively easy to calculate using widely available geospatial data and can be applied at spatial scales aligned with disturbance assessment and management. This study is one of the first attempts at measuring freshwater connectivity at broad scales and we expect our ability to do so will improve in the future as the underlying data improve (e.g., spatial and temporal resolution) and as new methods are developed to measure abundance and connectivity.

Implications for an integrated freshwater perspective

While the objectives of our analyses were to examine the distribution and diversity of freshwater abundance and connectivity and how landscape setting was related to these patterns, our freshwater clusters provide a framework in which connectivity attributes can be integrated into macroscale studies. Freshwater connectivity characteristics can have important implications for management and conservation because they play an integral role in physical, chemical, and biological integrity of freshwater ecosystems (Pringle 2001, 2003). Because of this, connectivity characteristics have been used as criteria to determine which water bodies are given protection under the U.S. Clean Water Act (Leibowitz et al. 2008). These connections can sustain hydrologic and biogeochemical conditions and support metapopulation dynamics. But, they also can impair ecosystems by transporting nutrients, contaminants, and spreading non-native species (Pringle 2001, Rahel 2007), which can have broad-scale regional consequences to downstream receiving water bodies (Freeman et al. 2007). Our integrated freshwater landscape clusters capture differences in lake, wetland, and stream connectivity attributes across a macroextent, and these clusters may also be related to differences in freshwater response such as nutrient transport processes or aquatic community composition. For example, landscape clusters with high proportions of lakes connected to upstream lakes (i.e., lake chains) may show a stronger spatial coherence in indices of water quality to regional changes in the landscape, or may have different species richness and diversity characteristics compared to clusters with less connected lake systems. This information could inform conservation and management strategies conducted at broad spatial scales. However, there is a need to evaluate the ecological relevancy of the integrated clusters by applying them to macroscale studies.

Further, there is growing recognition that upland-embedded and non-perennial systems, not just connected systems, can affect ecological integrity of downstream waters (U.S. EPA 2015). These systems may be considered isolated from persistent surface waters but can be connected to hydrologic networks through subsurface or overland flow and thus provide many of the ecosystem services that are associated with connected systems (Cohen et al. 2016, Rains et al. 2016). Without characterization of the distribution of geographically isolated systems at macroscales, we have poor understanding of the regional landscape context in which these systems operate, which can impede conservation and management actions to protect these vulnerable systems.

A macroscale perspective is warranted to study the patterns of the freshwater landscape at scales that are aligned with some of the leading disturbances to freshwater systems. Broad-scale disturbances such as land-use conversion and climate change threaten the integrity and function of the freshwater landscape by altering freshwater structural attributes. Urbanization and agricultural land use physically alter the size, shape, and connectivity characteristics of freshwater systems through water extraction and diversion, stream channelization, impoundment and burial of headwater bodies, wetland drainage, and altered flow regimes through dam and reservoir construction (Zedler and Kercher 2005, Freeman et al. 2007, Vörösmarty et al. 2010, Carpenter et al. 2011, Steele and Heffernan 2014, Van Meter and Basu 2015). In addition, changes in temperature and precipitation patterns associated with climate change are likely to impact freshwater systems and connectivity characteristics. Because the integrated freshwater landscape framework we developed here accounts for spatial variation in connectivity as well as density, it has the potential to contribute toward improved assessment of ecosystem integrity from disturbances and can help inform appropriate management actions necessary to maintain ecosystem services provided by our lakes, wetlands, and streams.

Acknowledgments

This project was supported by the National Science Foundation MacroSystems Biology Program in the Emerging Frontiers Division of Biological Sciences Directorate (EF-1065786). We thank Ed Bissell, Nicole Smith, and Scott Stopyak for their work in developing the LAGOS-NE_{GEO} database, the CSI-Limnology research group and Mary Bremigan for providing feedback throughout the development of this project, and the reviewers who provided insightful and constructive feedback. We thank Corinna Gries for helping upload data to the EDI depository and reviewing metadata documents. The manuscript idea was conceived of by CEF and refined and further developed through group meetings with the coauthors. CEF was the lead author drafting the majority of the text and creating tables, figures, and maps in the manuscript. All coauthors contributed to the conceptual development of the project, interpretation of results, and provided critical reviews and edits to the manuscript drafts. JFL substantially shaped the conceptual structure and approaches adopted in the study and performed PCA and *k*-means cluster analyses to quantify freshwater connectivity at landscape scales. SKO provided R code to perform random forest analyses and assisted with interpretation of model output. NKS assisted with wetland data acquisition, metadata development, and interpretation of wetland results. KSC provided constructive feedback and edits on conceptual structure of the manuscript. KEW contributed to the conceptual structure of the manuscript and assisted with developing R code for preprocessing steps and to create figures. PAS supervised and assisted in all stages of the development of the study by editing multiple drafts and providing constructive feedback on analyses and interpretation of results.

Literature Cited

Acuña, V., and K. Tockner. 2010. The effects of alterations in temperature and flow regime on organic carbon dynamics in Mediterranean river networks. Global Change Biology 16:2638–2650.

Ali, G. A., and A. G. Roy. 2009. Revisiting hydrologic sampling strategies for an accurate assessment of hydrologic connectivity in humid temperate systems. Geography Compass 3:350–374.

Aselmann, I., and P. J. Crutzen. 1989. Global distribution of natural freshwater wetlands and rice paddies, their net primary productivity, seasonality and possible methane emissions. Journal of Atmospheric Chemistry 8:307–358.

Bouvier, L. D., K. Cottenie, and S. E. Doka. 2009. Aquatic connectivity and fish metacommunities in wetlands of the lower Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences 66:933–948.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Butman, D., and P. A. Raymond. 2011. Significant efflux of carbon dioxide from streams and rivers in the United States. Nature Geoscience 4:839–842.

Butman, D., S. Stackpoole, E. Stets, C. P. McDonald, D. W. Clow, and R. G. Striegl. 2016. Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. Proceedings of the National Academy of Sciences USA 113:58–63.

Cardille, J. A., S. R. Carpenter, M. T. Coe, J. A. Foley, P. C. Hanson, M. G. Turner, and J. A. Vano. 2007. Carbon and water cycling in lake-rich landscapes: landscape connections, lake hydrology, and biogeochemistry. Journal of Geophysical Research: Biogeosciences 112:G02031.

Carpenter, S. R., E. H. Stanley, and M. J. V. Zanden. 2011. State of the world's freshwater ecosystems: physical, chemical, and biological changes. Annual Review of Environment and Resources 36:75–99.

Cohen, M. J., et al. 2016. Do geographically isolated wetlands influence landscape functions? Proceedings of the National Academy of Sciences USA 1978–1986:201512650. https://doi.org/10.1073/pnas.1512650113

Cowardin, L. M., V. Carter, F. C. Golet, and E. T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C., USA.

Crump, B. C., H. E. Adams, J. E. Hobbie, and G. W. Kling. 2007. Biogeography of bacterioplankton in lakes and streams of an Arctic tundra catchment. Ecology 88:1365–1378.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Dahl, T. E. 2006. Status and trends of wetlands in the conterminous United States 1998 to 2004. U.S. Department of the Interior; Fish and Wildlife Service, Washington, D.C., USA.

Datry, T., S. T. Larned, and K. Tockner. 2014. Intermittent rivers: a challenge for freshwater ecology. BioScience 64:bit027.

Detenbeck, N. E., C. A. Johnston, and G. J. Niemi. 1993. Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area. Landscape Ecology 8:39–61.

Downing, J. A., J. J. Cole, C. A. Duarte, J. J. Middelburg, J. M. Melack, Y. T. Prairie, P. Kortelainen, R. G. Striegl, W. H. McDowell, and L. J. Tranvik. 2012. Global abundance and size distribution of streams and rivers. Inland Waters 2:229–236.

Downing, J. A., and C. M. Duarte. 2009. Abundance and size distribution of lakes, ponds, and impoundments. Elsevier, Oxford, UK.

Downing, J. A., et al. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnology and Oceanography 51:2388–2397.

Fergus, C., J. Lapierre, S. Oliver, N. Skaff, K. Cheruvelil, K. Webster, C. Scott, and P. Soranno. 2017*a*. Freshwater connectivity clusters for lakes, wetlands, and streams at the Hydrologic Unit 12 scale in the Midwest and Northeast U.S.A. – freshwater metric variables and k-means cluster assignment. Environmental Data Initiative. https://doi.org/10.6073/pasta/1dc5b024390539f1f7dda70470d35766

Fergus, C., J. Lapierre, S. Oliver, N. Skaff, K. Cheruvelil, K. Webster, C. Scott, and P. Soranno. 2017*b*. Integrated freshwater abundance and connectivity clusters at the Hydrologic Unit 8 scale for the Midwest and Northeast U.S.A. – freshwater metric variables and k-means cluster assignment. Environmental Data Initiative. https://doi.org/10.6073/pasta/e69b7495674e403baa19a22ffbfb40e1

Fisher, S. G., R. A. Sponseller, and J. B. Heffernan. 2004. Horizons in stream biogeochemistry: flowpaths to progress. Ecology 85:2369–2379.

Freeman, M. C., C. M. Pringle, and C. R. Jackson. 2007. Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales. JAWRA Journal of the American Water Resources Association 43:5–14.

Fritz, K. M., E. Hagenbuch, E. D'Amico, M. Reif, P. J. Wigington, S. G. Leibowitz, R. L. Comeleo, J. L. Ebersole, and T.-L. Nadeau. 2013. Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. JAWRA Journal of the American Water Resources Association 49:867–882.

Genuer, R., J. M. Poggi, and C. Tuleau-Malot. 2016. VSURF: Variable selection using random forests. R package version 1.0.3. https://CRAN.R-project.org/package=VSURF

Heffernan, J. B., et al. 2014. Macrosystems ecology: understanding ecological patterns and processes at continental scales. Frontiers in Ecology and the Environment 12:5–14.

Humborg, C., C.-M. Mörth, M. Sundbom, H. Borg, T. Blenckner, R. Giesler, and V. Ittekkot. 2010. CO2 supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. Global Change Biology 16:1966–1978.

Jones, J. A. 2011. Hydrologic responses to climate change: considering geographic context and alternative hypotheses. Hydrological Processes 25: 1996–2000.

Jones, J. A., et al. 2012. Ecosystem processes and human influences regulate streamflow response to climate change at long-term ecological research sites. BioScience 62:390–404.

Kling, G. W., G. W. Kipphut, M. M. Miller, and W. J. O'Brien. 2000. Integration of lakes and streams in a landscape perspective: the importance of material processing on spatial patterns and temporal coherence. Freshwater Biology 43:477–497.

Kratz, T., K. Webster, C. Bowser, J. Maguson, and B. Benson. 1997. The influence of landscape position on lakes in northern Wisconsin. Freshwater Biology 37:209–217.

Larned, S. T., T. Datry, D. B. Arscott, and K. Tockner. 2010. Emerging concepts in temporary-river ecology. Freshwater Biology 55:717–738.

Lehner, B., and P. Döll. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296:1–22.

Leibowitz, S. G., P. J. Wigington, M. C. Rains, and D. M. Downing. 2008. Non-navigable streams and adjacent wetlands: addressing science needs following the Supreme Court's Rapanos decision. Frontiers in Ecology and the Environment 6:364–371.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Lohse, K. A., P. D. Brooks, J. C. McIntosh, T. Meixner, and T. E. Huxman. 2009. Interactions between biogeochemistry and hydrologic systems. Annual Review of Environment and Resources 34:65–96.

Lottig, N. R., E. H. Stanley, P. C. Hanson, and T. K. Kratz. 2011. Comparison of regional stream and lake chemistry: differences, similarities, and potential drivers. Limnology and Oceanography 56: 1551–1562.

Martin, S. L., and P. A. Soranno. 2006. Lake landscape position: relationships to hydrologic connectivity and landscape features. Limnology and Oceanography 51:801–814.

McDonald, C. P., J. A. Rover, E. G. Stets, and R. G. Striegl. 2012. The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. Limnology and Oceanography 57:597–606.

Meybeck, M. 1995. Global distribution of lakes. Pages 1–35 *in* P. A. Lerman, P. D. D. M. Imboden, and P. J. R. Gat, editors. Physics and chemistry of lakes. Springer, Berlin, Germany; Heidelberg, Germany.

Moore, J. W. 2015. Bidirectional connectivity in rivers and implications for watershed stability and management. Canadian Journal of Fisheries and Aquatic Sciences 72:785–795.

Müller, R. A., M. N. Futter, S. Sobek, J. Nisell, K. Bishop, and G. A. Weyhenmeyer. 2013. Water renewal along the aquatic continuum offsets cumulative retention by lakes: implications for the character of organic carbon in boreal lakes. Aquatic Sciences 75:535–545.

Mushet, D. M., A. J. K. Calhoun, L. C. Alexander, M. J. Cohen, E. S. DeKeyser, L. Fowler, C. R. Lane, M. W. Lang, M. C. Rains, and S. C. Walls. 2015. Geographically isolated wetlands: rethinking a misnomer. Wetlands 35:423–431.

Nadeau, T.-L., and M. C. Rains. 2007. Hydrological connectivity between headwater streams and downstream waters: how science can inform policy. JAWRA Journal of the American Water Resources Association 43:118–133.

Nelson, C. E., S. Sadro, and J. M. Melack. 2009. Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. Limnology and Oceanography 54:1292–1305.

Oksanen, J., et al. 2017. vegan: community ecology package. https://CRAN.R-project.org/package=vegan

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540: 418–422.

Pringle, C. M. 2001. Hydrologic connectivity and the management of biological reserves: a global perspective. Ecological Applications 11:981–998.

Pringle, C. 2003. What is hydrologic connectivity and why is it ecologically important? Hydrological Processes 17:2685–2689.

Quinlan, R., A. M. Paterson, R. I. Hall, P. J. Dillon, A. N. Wilkinson, B. F. Cumming, M. S. V. Douglas, and J. P. Smol. 2003. A landscape approach to examining spatial patterns of limnological variables and long-term environmental change in a southern Canadian lake district. Freshwater Biology 48:1676–1697.

Racchetti, E., M. Bartoli, E. Soana, D. Longhi, R. R. Christian, M. Pinardi, and P. Viaroli. 2010. Influence of hydrological connectivity of riverine wetlands on nitrogen removal via denitrification. Biogeochemistry 103:335–354.

Rahel, F. J. 2007. Biogeographic barriers, connectivity and homogenization of freshwater faunas: It's a small world after all. Freshwater Biology 52:696–710.

Rains, M. C., S. G. Leibowitz, M. J. Cohen, I. F. Creed, H. E. Golden, J. W. Jawitz, P. Kalla, C. R. Lane, M. W. Lang, and D. L. McLaughlin. 2016. Geographically isolated wetlands are part of the hydrological landscape. Hydrological Processes 30:153–160.

Raymond, P. A., et al. 2013. Global carbon dioxide emissions from inland waters. Nature 503:355–359.

Read, E. K., et al. 2015. The importance of lake-specific characteristics for water quality across the continental United States. Ecological Applications 25: 943–955.

Riera, J. L., J. J. Magnuson, T. K. Kratz, and K. E. Webster. 2000. A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. Freshwater Biology 43:301–318.

Rodriguez-Iturbe, I., M. Marani, R. Rigon, and A. Rinaldo. 1994. Self-organized river basin landscapes: fractal and multifractal characteristics. Water Resources Research 30:3531–3539.

Sadro, S., C. E. Nelson, and J. M. Melack. 2011. The influence of landscape position and catchment characteristics on aquatic biogeochemistry in high-elevation lake-chains. Ecosystems 15:363–386.

Seaber, P. R., F. P. Kapinos, and G. L. Knapp. 1987. Hydrologic unit maps. U.S. Government Printing Office, Washington, D.C., USA.

Smith, S. V., W. H. Renwick, J. D. Bartley, and R. W. Buddemeier. 2002. Distribution and significance of small, artificial water bodies across the United States landscape. Science of the Total Environment 299:21–36.

Soranno, P. A., K. S. Cheruvelil, T. Wagner, K. E. Webster, and M. T. Bremigan. 2015*a*. Effects of land use on lake nutrients: the importance of scale, hydrologic connectivity, and region. PLoS ONE 10:e0135454.

Soranno, P. A., et al. 1999. Spatial variation among lakes within landscapes: ecological organization along lake chains. Ecosystems 2:395–410.

Soranno, P. A., et al. 2015*b*. Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. GigaScience 4:28.

Steele, M. K., and J. B. Heffernan. 2014. Morphological characteristics of urban water bodies: mechanisms of change and implications for ecosystem function. Ecological Applications 24:1070–1084.

Strahler, A. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 8:913–920.

Strayer, D. L., R. E. Beighley, L. C. Thompson, S. Brooks, C. Nilsson, G. Pinay, and R. J. Naiman. 2003. Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. Ecosystems 6:407–423.

Tiner, R. W. 2003. Geographically isolated wetlands of the United States. Wetlands 23:494–516.

U.S. EPA. 2015. Connectivity of Streams and Wetlands to Downstream Waters: A Review and Synthesis of the Scientific Evidence. Final Report. U.S. Environmental Protection Agency, Washington, D.C., USA.

Van Meter, K. J., and N. B. Basu. 2015. Signatures of human impact: size distributions and spatial organization of wetlands in the Prairie Pothole landscape. Ecological Applications 25:451–465.

Verpoorter, C., T. Kutser, D. A. Seekell, and L. J. Tranvik. 2014. A global inventory of lakes based on high-resolution satellite imagery. Geophysical Research Letters 41:2014GL060641.

Vörösmarty, C. J., et al. 2010. Global threats to human water security and river biodiversity. Nature 467:555–561.

Webster, K. E., P. A. Soranno, K. S. Cheruvelil, M. T. Bremigan, J. A. Downing, P. D. Vaux, T. R. Asplund, L. C. Bacon, and J. Connor. 2008. An empirical evaluation of the nutrient-color paradigm for lakes. Limnology and Oceanography 53:1137–1148.

Weller, C. M., M. C. Watzin, and D. Wang. 1996. Role of wetlands in reducing phosphorus loading to surface water in eight watersheds in the Lake

Champlain Basin. Environmental Management 20: 731–739.
Winter, T. C. 2000. The vulnerability of wetlands to climate change: a hydrologic landscape perspective. JAWRA Journal of the American Water Resources Association 36:305–311.
Winter, T. C. 2001. The concept of hydrologic landscapes. JAWRA Journal of the American Water Resources Association 37:335–349.
Wolock, D. M., T. C. Winter, and G. McMahon. 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environmental Management 34:S71–S88.
Wright, C. K., and M. C. Wimberly. 2013. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proceedings of the National Academy of Sciences USA 110:4134–4139.
Zedler, J. B., and S. Kercher. 2005. Wetland resources: status, trends, ecosystem services, and restorability. Annual Review of Environment and Resources 30:39–74.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1911/full