

Relative importance of environmental, geographic, and spatial variables on zooplankton metacommunities

TAD DALLAS† AND JOHN M. DRAKE

Odum School of Ecology, University of Georgia, Athens, Georgia 30606 USA

Citation: Dallas, T., and J. M. Drake. 2014. Relative importance of environmental, geographic, and spatial variables on zooplankton metacommunities. *Ecosphere* 5(9):104. <http://dx.doi.org/10.1890/ES14-00071.1>

Abstract. Understanding the factors responsible for structuring ecological communities is a central goal in community ecology. Previous work has focused on determining the relative roles of two classes of variables (e.g., spatial and environmental) on community composition. However, this approach may ignore the disproportionate impact of variables within classes, and is often confounded by spatial autocorrelation leading to collinearity among variables of different classes. Here, we combine pattern-based metacommunity and machine learning analyses to characterize metacommunity structure of zooplankton from lakes in the northeast United States and to identify environmental, spatial, and geographic covariates associated with metacommunity structure. Analyses were performed for the entire metacommunity and for three zooplankton subsets (cladocerans, copepods, and rotifers), as the variables associated with community structure in these groups were hypothesized to differ. Species distributions of all subsets adhered to an environmental, spatial, and/or geographic gradient, but differed in metacommunity pattern, as copepod species distributions responded independently of one another, while the entire zooplankton metacommunity, cladocerans, and rotifers replaced one another in discrete groups. While environmental variables were nearly always the most important to metacommunity structure, the relative importance of variables differed among zooplankton subsets, suggesting that zooplankton subsets differ in their environmental tolerances and dispersal-limitation.

Key words: boosted regression tree; community assembly; Eastern Lake Survey; Elements of Metacommunity Structure; metacommunity; zooplankton.

Received 4 March 2014; **revised** 7 May 2014; **accepted** 9 May 2014; **final version received** 16 June 2014; **published** 19 September 2014. Corresponding Editor: J. P. Benstead.

Copyright: © 2014 Dallas and Drake. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. <http://creativecommons.org/licenses/by/3.0/>

† **E-mail:** tdallas@uga.edu

INTRODUCTION

Identifying the mechanisms underlying species distributions and community structure is a central focus of community ecology (Hairston et al. 1960, Soininen et al. 2007, De Meester 2011). Two distinct explanations of community structure have emerged: niche assembly and dispersal assembly. According to niche assembly, community composition is controlled predominantly by environmental forces, emphasizing the physiological tolerances of species to environmental

conditions. By contrast, dispersal assembly holds that community composition is limited by the ability of species to reach new habitats, making spatial (e.g., distance among sites) or geographic variables (e.g., patch size, elevation, etc.) stronger determinants of community composition than environmental variables (Weiher et al. 2011). This makes the assumption that variables relating to patch size or distance among patches are directly related to dispersal likelihood. Past studies seeking to quantify the relative importance of environmental and spatial factors have obtained

conflicting results (Cornell and Lawton 1992, Pinel-Alloul et al. 1995, Cottenie 2005), possibly because actual community assembly occurs through a combination of niche and dispersal processes (Mouillot 2007) or as a result of variable choice or multicollinearity among predictors. Methods capable of accounting for collinearity among variable sets are therefore important in understanding the structure of ecological communities.

Difficulties distinguishing factors responsible for structuring ecological communities have been a major impediment to our understanding of community assembly, suggesting the need for alternative analytical approaches to metacommunity analysis. Previous methods have argued that mechanistic metacommunity models (i.e., species sorting, neutral dynamics, mass effects, or patch dynamics; Leibold et al. 2004, Holyoak et al. 2005) can be distinguished through the partitioning of spatial and environmental variables, using the significance of the variance partitions of environment and spatial variables to be indicative of a metacommunity model. Variance partitioning approaches may be insensitive to autocorrelation among variable sets, and are not typically used to analyze variance explained from single variables, but from entire classes of variables (Beisner et al. 2006). This may obscure the relative roles of variable classes, since one strongly associated variable could drive the explained variance for an entire variable class. As a result of these issues with the application of variance partitioning to the study of community composition, we do not attempt to distinguish among different metacommunity models (e.g., neutral dynamics), which are traditionally discerned using a variance partitioning framework (Cottenie 2005).

Instead, we combine two analyses to determine how a metacommunity is structured, and what variables are most associated with metacommunity structure in lake ecosystems. Lake ecosystems offer an excellent opportunity for studying metacommunity structure, as each lake offers a bounded community with a set of unambiguously measurable environmental attributes (e.g., depth, pH). Community ecologists have recognized that zooplankton communities therefore represent an ideal study system for examining the drivers of community structure,

leading to a rich body of literature (Cottenie et al. 2003, Havel and Shurin 2004, Cottenie and De Meester 2005, Medley and Havel 2007). Despite the amount of attention paid to these tractable communities, conflicting reports exist on whether geographic factors or environmental variables (Cottenie 2005) are more important in determining community composition.

To address this conflict, we used the Elements of Metacommunity Structure (EMS) framework of Leibold and Mikkelsen (2002) together with regression tree analysis to determine drivers of community structure and species distributions of zooplankton of the Eastern United States. The EMS framework employs three statistics that, when taken together, can be used to distinguish among metacommunity structures (Leibold and Mikkelsen 2002) (Fig. 1). The three statistics that compose the EMS framework are coherence, turnover, and boundary clumping (Table 1). Coherence is the number of embedded absences in a species range, turnover is the number of times species replace one another across their respective ranges (similar to the C-score in co-occurrence analysis), and boundary clumping is the tendency of species ranges to clump, or form modules. These statistics together allow for the classification of metacommunities as random, checkerboard, or adhering to a structuring gradient, with species responding independently (Gleasonian metacommunity, named for botanist H.A. Gleason; Gleason 1926) or as discrete groups of species (Clementsian metacommunity, named for ecologist Frederic Clements; Clements 1916). This approach allows for the classification of metacommunities into “patterns” or “types”, but also provides a quantitative gradient (based on reciprocal averaging) along which the metacommunity is structured. This gradient can then be related to environmental, geographic, and spatial variables. Here, we use boosted regression trees (BRT) (De’Ath 2007, Elith et al. 2008) to relate the gradient along which communities are structured to environmental, geographic, and spatial variables. Boosted regression trees are flexible to nonlinearities and collinearity inherent in multivariate community data, in which environment and geography are intrinsically linked.

In this article, we analyze data on zooplankton species distributions among 139 lakes in a 4×10^5 km² area of the northeast United States (US EPA

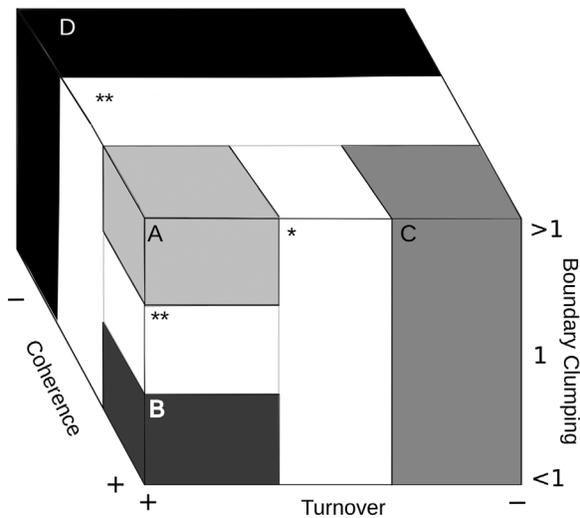


Fig. 1. The three-dimensional space created by the three statistics used to determine metacommunity structure (taken from Dallas 2014). Quasi-structures proposed by Presley et al. (2010) (marked with a single asterisk) and metacommunity structures unable to be determined using the EMS analysis (*Gleasonian* and *random*; double asterisk) are not considered in our current framework. Instead, the EMS analysis can distinguish among *Clementsian* (A), *evenly spaced gradients* (B), *nested subsets* (C), and *checkerboard* (D) structures.

1990) to determine the relative influence of a suite of environmental, geographic, and spatial variables on zooplankton species distributions. These data have been used for at least one other

study on zooplankton community assembly and structure (Leibold et al. 2010). However, previous studies have used “problematic” methods (see Anderson et al. 2011 for a discussion of inferring metacommunity dynamics from variance partitioning), and focused on the impact of entire classes of variables (e.g., spatial variables versus environmental variables) rather than considering the individual impact of each variable relative to other variables. Further, previous work has largely focused on the zooplankton community as a whole (Cottenie et al. 2003), or focused on a particular zooplankton subset (e.g., rotifers; Fontaneto et al. 2006), but see Leibold et al. (2010). However, zooplankton groups likely differ in their responses to environmental and geographic variables, and dispersal capability (Havel and Shurin 2004). To address these issues, we utilize the EMS analysis coupled with boosted regression trees, to identify variables most important to community structure of the entire zooplankton metacommunity and taxonomic subsets (e.g., rotifers), and to quantify the relative importance of these variables to community structure. Variables examined span three classes, including environmental (e.g., water chemistry variables), spatial (e.g., distance among sites), and geographic (e.g., elevation, lake size) variables. We hypothesized that zooplankton subsets would differ in which variables were important to community structure and the relative importance of variables relative to others. Further, given the dispersal capabilities of many zooplankton taxa (Fontaneto et al. 2006), we

Table 1. Idealized metacommunity structures and associated significance signs for the three attributes used to identify the best fit pattern. Significance signs are based on the position of the calculated statistic relative to the null distribution of the statistic, with “–” indicating a statistic less than the null distribution, “ns” indicating no significant difference between empirical statistic and null statistic distribution, and “+” indicating a statistic larger than the null distribution. These attributes form the axes of a three-dimensional space depicted in Fig. 1. An asterisk (*) indicates metacommunity patterns determined by asserting the null hypothesis, and therefore, the EMS analysis may not be able to distinguish these metacommunity types.

Pattern	Definition	Coherence	Turnover	Boundary clumping
Random*	Species ranges follow no detectable gradient	ns	+; –; ns	+; –; ns
Checkerboard	Species pairs have mutually exclusive distributions	–	+; –; ns	+; –; ns
Nested subsets	Species ranges form nested groups	+	–	+; –; ns
Evenly spaced gradients	Species ranges distributed evenly across gradient	+	+	–
Gleasonian*	Species ranges adhere to gradient, but do so individually	+	+	ns
Clementsian	Species ranges form groups, which replace each other along gradient	+	+	+

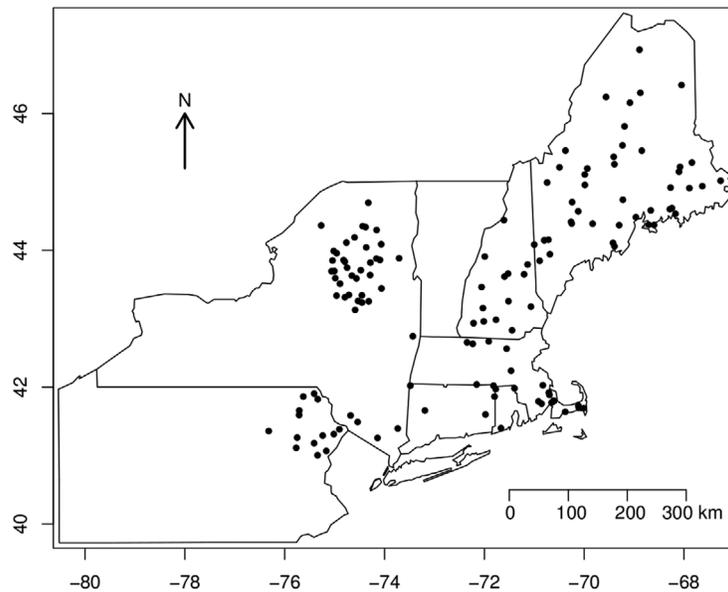


Fig. 2. Geographic locations of lakes sampled by the EPA from 1984 to 1986 and used in the present analyses.

hypothesized that environmental and geographic variables would play a larger role than spatial distances among sites.

METHODS

Data collection

Between 1984 and 1986, the United States Environment Protection Agency conducted a National Surface Water Survey in an effort to document the biotic status of lakes in the United States, specifically those that may be sensitive to acidification. Here, we examined water chemistry and zooplankton community data sampled during the summer of 1986 as part of the Eastern Lake Survey (Fig. 2), a smaller component of the National Surface Water Survey (US EPA 1990). Lakes were chosen using a systematic random sample (Herlihy et al. 1991) from a subset of lakes with acid neutralizing capacity (ANC) $<400 \mu\text{g/L}$, depth $>1.5 \text{ m}$, and low nitrogen or phosphorous concentrations. Water chemistry variables were measured following the methods of Mitchell-Hall et al. (1989) from a sample taken just after lake turnover near the deepest part of the sampled lake with a 6.2 L Van Dorn acrylic bottle filled from a depth of 1.5 m. Chlorophyll *a* was determined spectrophotometrically following the EPA Acidic Deposition Analytical Methods Man-

ual (US EPA 1987). Zooplankton were sampled using an 80 μm mesh Wisconsin bucket net for three vertical tows of the entire water column (Tessier and Horwitz 1991). To limit the influence of transient species, a species must have been documented in more than one of the three plankton tows to be considered present in the lake. Further, species that occurred in only one lake were removed from analysis, as they can exert strong influences on coherence and boundary clumping, potentially biasing results (Presley et al. 2009, Keith et al. 2011).

Variable selection

Variables were selected a priori but informed by previous studies on metacommunity structure (Cottenie 2005, Leibold et al. 2010). Given the breadth of variables measured in the EPA sampling effort, we selected ten environmental variables, seven geographic variables, and two spatial variables for our analyses (Table 2). Spatial variables were the first two vectors from a principal coordinates analysis (PCoA) on the distance matrix generated by determining the geodesic distance between all sites (determined using R package *ape*; Paradis et al. 2004). These two vectors constituted 98.9% of the eigenvalues for the distance matrix. Environmental variables were all related to patch quality, while geograph-

Table 2. Summary statistics for the environmental and geographical variables from 139 lakes in the northeast United States used in the boosted regression tree analysis.

Variable	Units	Mean (SE)	Min	Max
Environmental				
Chlorophyll <i>a</i>	$\mu\text{g L}^{-1}$	4.99 (0.08)	0.35	132.30
Nitrate	mg L^{-1}	2.09 (0.34)	0.00	18.21
Ammonium	mg L^{-1}	2.15 (0.92)	0.00	25.50
Phosphorous	$\mu\text{g L}^{-1}$	9.64 (11.35)	0.00	90.00
DIC	mg L^{-1}	1.04 (0.09)	0.05	4.35
DOC	mg L^{-1}	4.08 (0.45)	0.40	12.42
Carbonate	$\mu\text{g L}^{-1}$	0.12 (0.00)	0.00	1.55
Potassium	$\mu\text{g L}^{-1}$	11.42 (4.40)	1.56	42.70
Turbidity	Secchi depth (cm)	72.66 (3.08)	13.00	157.00
pH	...	6.60 (0.08)	4.46	8.01
Runoff	m/year	0.55 (0.00)	0.25	0.76
Geographic				
Elevation	m	295.70 (18.31)	1.50	791.00
Lake size	ha	381.40 (20.95)	3.00	790.00
Inlets	no.	0.96 (0.17)	0.00	8.00
Lake type
Watershed area	ha	3086.02 (1.1e7)	13.00	81424.00
Lake volume	10^6 m^3	475.20 (27.17)	31.00	1075.00
Spatial				
PCoA vector 1	...	1.56 e-17 (0.01)	-1.91	1.99
PCoA vector 2	...	2.07e-17 (0.01)	-1.86	2.15

ic variables were related to the physical aspects of lakes. For instance, lake size, watershed area, and the number of inlets all impact the potential for zooplankton to move and/or successfully colonize habitats. Therefore, physical variables may relate to dispersal likelihood, but they are also likely related to environmental parameters.

Elements of metacommunity structure analysis

Community matrices (site-by-species) of presence-absence data were assembled for each of three major groups; rotifers (44 spp.), cladocerans (29 spp.), and copepods (20 spp.). Analyses were performed on each group separately and on the community as a whole (93 total spp.). Within a taxonomic group, lakes containing no species occurrences were removed prior to analysis. However, this occurred infrequently (one and two lakes for rotifers and copepods, respectively) and is unlikely to influence our overall conclusions.

The EMS framework uses presence-absence data to distinguish metacommunity pattern by calculating three metrics; coherence, species turnover, and range boundary clumping (Leibold and Mikkelsen 2002, Presley et al. 2010), which form a three-dimensional space within which different metacommunity patterns occupy different regions (Table 1, Fig. 1). Matrices were first

ordinated via reciprocal averaging (Gauch 1982), a technique that arranges sites with most similar community composition and species with most similar distributions closer together. This ordination differs slightly from other ordination techniques, such as principal components analysis, in that higher values (species presences in our study) are concentrated along the matrix diagonal (Gauch 1982). The weights obtained from reciprocal averaging represent a gradient along which species distributions are structured, provided that species distributions are not *random* or *checkerboard*. Coherence was measured by counting the number of embedded absences in the ordinated matrix and compared against the number of embedded absences from 1000 ordinated null matrices. Significance was determined by comparing observed versus null embedded absences using a z-test. Negative coherence (more embedded absences than expected under null model) indicates a checkerboard pattern, in which species occurrences among sites are mutually exclusive. Positive coherence indicates that species ranges have fewer embedded absences than expected under null model simulations. To qualify for further analysis, community matrices must exhibit positive coherence (i.e., fewer embedded absences than null model simulations).

Turnover was quantified by calculating the number of times one species replaced another between sites, after species distributions are made completely coherent. Therefore, this does not include gaps in species ranges in the turnover metric; only instances where species replace one another from site to site are considered. This metric is compared to the distribution of turnover values obtained through 1000 ordinated null simulations using a z-test. Low turnover is indicative of nested subsets, whereas significantly high turnover means that species replace one another at the ends of their ranges more often than expected by chance. This indicates that the metacommunity can be structured as Clementsian, Gleasonian, or have evenly spaced gradients (Table 1; Leibold and Mikkelsen 2002). Boundary clumping was quantified using the Morisita's index, a measure of the dispersion of species occurrences among sites (Morisita 1971). A Morisita's index (I) of one indicates boundaries are not clumped, while values greater than one ($I > 1.0$) or less than one ($I < 1.0$) indicate clumped or hyperdispersed boundaries, respectively. Statistical significance of the Morisita's index was determined using a chi-squared test.

It is important to note that our interpretation of the EMS analysis differs slightly from that of Leibold and Mikkelsen (2002) and Presley et al. (2010). In prior studies, the non-significance of coherence (no difference in embedded absences between calculated statistic and null distribution) was interpreted as evidence that the community was randomly structured. We believe this is to mistakenly assert the null hypothesis. That is, we assert that failure to detect an effect indicates that the test is unable to distinguish positive from negative coherence, and does not provide evidence that the community is actually random. Therefore, to interpret a non-significant result as evidence for random metacommunity structure is to commit the error of accepting the null hypothesis. Accordingly, we do not consider two patterns (Fig. 1) (random and Gleasonian) that are part of the original formulation of Leibold and Mikkelsen (2002) to be discernible through the EMS analysis (Dallas 2014).

The choice of null model is not unambiguous, as many permutation algorithms are available, each differing in their type I and type II error properties (Gotelli 2000). In our study, observed

species richness per site was fixed (row totals fixed), as lakes differ in their suitability and richness may be contingent upon differences in this suitability. Zooplankton species distributions may be dispersal-limited at the geographic scale of the current study (Shurin et al. 2009, De Meester 2011). To incorporate this into the null model, species occurrences at sites were determined using the marginal frequencies in the observed community matrices as probabilities of occurrence in the null matrices (Wright et al. 1997). In addition to biological realism, this null model algorithm (fixed row – proportional column) has desirable type I and II error properties relative to most others (Gotelli and Graves 1996, Gotelli 2000).

Boosted regression tree analysis

The EMS analysis provides insight into metacommunity patterns, but not into the variables potentially responsible for creating the observed pattern. Correlations with the values obtained from reciprocal averaging can provide some information (Presley and Willig 2009, Keith et al. 2011), but don't provide information about the relative influence of individual variables when considering a suite of potentially interrelated factors. Regression tree analysis is a useful tool for prediction and identification of relevant predictor variables, but has only recently been applied to ecological systems (De'Ath 2007, Elith et al. 2008). Specifically, boosted regression trees (BRT) bypass many of the issues with traditional approaches, such as collinearity of predictor variables, nonlinear relationships between predictors and response (Elith et al. 2008), or the fallibility of commonly applied approaches like variance partitioning (Gilbert and Bennett 2010).

“Boosting” refers to the process whereby many trees are created in order to extract general “weak” rules, which are then combined to enhance predictive ability. The optimal number of trees was determined using k -fold cross-validation ($k = 10$) to avoid overfitting. The learning rate ($l = 0.001$), or shrinkage parameter, determines the degree to which each new tree contributes to the overall model. An interaction depth of 2 was used in order to allow for two-way variable interactions, thereby reducing the effect of collinearity. Cross validation was also used to determine the optimal number of trees

Table 3. Elements of metacommunity structure analysis for primary and secondary ordination axes. Coherence was calculated by determining the number of embedded absences in the interaction matrix (Abs) and relating this to a null distribution, with mean and standard deviation reported (Mean (SD)). The number of replacements (Rep) and the mean and standard deviation of the turnover statistic (Mean (SD)) are divided by $10e^3$. Range clumping was determined by calculating Morisita's index (I).

Group	Coherence			Turnover			Clumping	
	Abs	Mean (SD)	p	Rep	Mean (SD)	p	I	p
Primary axis								
Entire	8135	9813 (110)	<0.0001	154	114 (15)	0.0084	1.92	<0.0001
Cladocera	1688	2511 (75)	<0.0001	21	13 (3)	0.0037	1.59	0.0377
Copepod	985	1532 (63)	<0.0001	8	4 (1)	<0.0001	1.82	0.0517
Rotifer	3254	4138 (79)	<0.0001	27	14 (3)	<0.0001	2.37	<0.0001
Secondary axis								
Entire	7669	9850 (106)	<0.0001	156	84 (12)	<0.0001	2.34	<0.0001
Cladocera	1599	2554 (77)	<0.0001	21	11 (2)	<0.0001	4.06	<0.0001
Copepod	1252	1585 (67)	<0.0001	11	6 (2)	0.0038	0.86	0.3942
Rotifer	3313	4185 (81)	<0.0001	36	20 (4)	<0.0001	2.79	<0.0001

(upper limit = 20000 trees) before model performance declined. The relative contribution (RC) of each predictor variable was determined by randomly permuting each predictor variable and quantifying the reduction in model performance, a method that is free of classical assumptions about normality and equal variance (Anderson 2001). Relative contribution estimates were then based on the number of times a given predictor variable was selected for splitting, weighted by the degree the split improves model performance. This metric was averaged across all trees built in the model, and scaled between 0 (predictor had no contribution) to 100 (predictor is very important).

The EMS analysis was performed using the *metacom* package (Dallas 2013) in R version 2.15.1 (R Development Core Team 2013), which relies on functions from the *vegan* package (Oksanen et al. 2012). A significance level of $\alpha = 0.05$ was used for all elements of metacommunity structure. BRT models were fit using R package *gbm* (Ridgeway 2012).

RESULTS

Metacommunity structure

Zooplankton metacommunity and subcommunity patterns, determined through the analysis of coherence, range turnover and boundary clumping, remained similarly structured regardless of ordination axis (Table 3). The entire zooplankton metacommunity, along with cladoceran and rotifer metacommunities, exhibited a Clement-

sian structure, indicating that groups of species responded to the gradient rather than each species responding independently (Table 3, Fig. 3). Copepods, by contrast, adhered to what has been previously called a *Gleasonian* structure along both ordination axes. We interpret this result to imply that the copepod metacommunity is structured along a gradient representing environment, space, and geography, but that we are unable to discern how species replace one another along the structuring gradient, apart from the result that species replaced one another more often than expected by chance. Concretely, we are unable to distinguish the copepod metacommunity as Gleasonian, but recognize that species ranges are responding to a gradient in an organized fashion.

Structuring mechanisms

Metacommunities differed in the relative contributions (RC) of predictor variables to the primary axis scores obtained by reciprocal averaging (Fig. 4). The top three variables related to the gradient structuring the entire zooplankton metacommunity included a measure of spatial distance (i.e., second PCoA vector) (RC = 24.63), pH (RC = 16.46), and chlorophyll *a* (RC = 8.70). Despite the importance of the two variables representing spatial distance among sites, environmental variables were most important to the structure of the entire zooplankton metacommunity ($RC_E = 60.03$). Geographic variables were largely unimportant, accounting for between 11% and 18% of the relative contribution

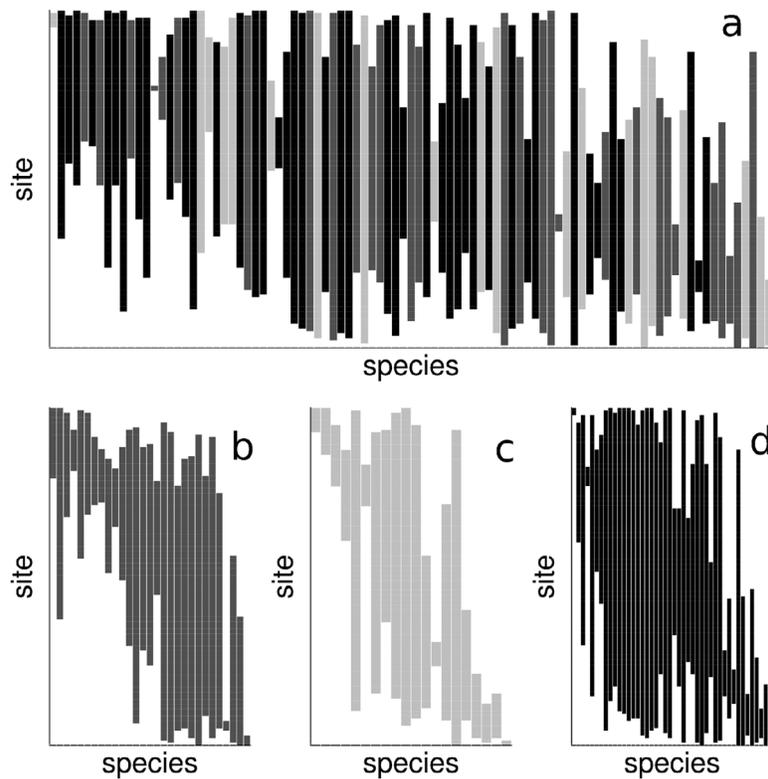


Fig. 3. Zooplankton species distributions along a structuring gradient (i.e., the dominant ordination axis from reciprocal averaging). Species ranges across sites are made entirely coherent (no gaps in a species range) for visualization. The entire zooplankton community is shown in (A) with each taxonomic subset shaded to match subset species distribution plots for copepods (B), cladocerans (C), and rotifers (D).

values. Environmental variables had the highest summed relative contribution values, even when considering geographic variables to be proxies for dispersal likelihood. This is true for all metacommunities apart from copepods, in which spatial variables ($RC_5 = 45.37$) played a substantial role.

Aside from spatial distances among sites, the copepod metacommunity was structured by phosphorus ($RC = 10.27$), chlorophyll *a* ($RC = 11.36$), and turbidity ($RC = 7.46$). The cladoceran metacommunity was structured largely by resource availability, as chlorophyll *a* exerted the strongest effect ($RC = 25.05$), followed by dissolved organic carbon (DOC) ($RC = 11.89$) and spatial distance among sites ($RC = 17.22$). Rotifer communities were structured predominantly by pH ($RC = 42.68$), dissolved inorganic carbon (DIC) ($RC = 15.11$), and spatial distance among sites ($RC = 10.09$).

DISCUSSION

The zooplankton metacommunity and all subsets were significantly structured along a gradient. The entire zooplankton, cladoceran, and rotifer metacommunities were Clementsian, as species ranges adhered to a gradient, and formed discrete groups that replaced one another along the gradient. We were unable to determine the pattern that the copepod metacommunity adhered to as a result of non-significant boundary clumping corresponding to a small or null effect, though the metacommunity was structured along a gradient as indicated by significantly coherent species ranges, and higher turnover than expected under our null model conditions. The gradient responsible for structuring zooplankton communities was largely environmental, which supports the *niche-assembly* paradigm. However, we recognize that spatial

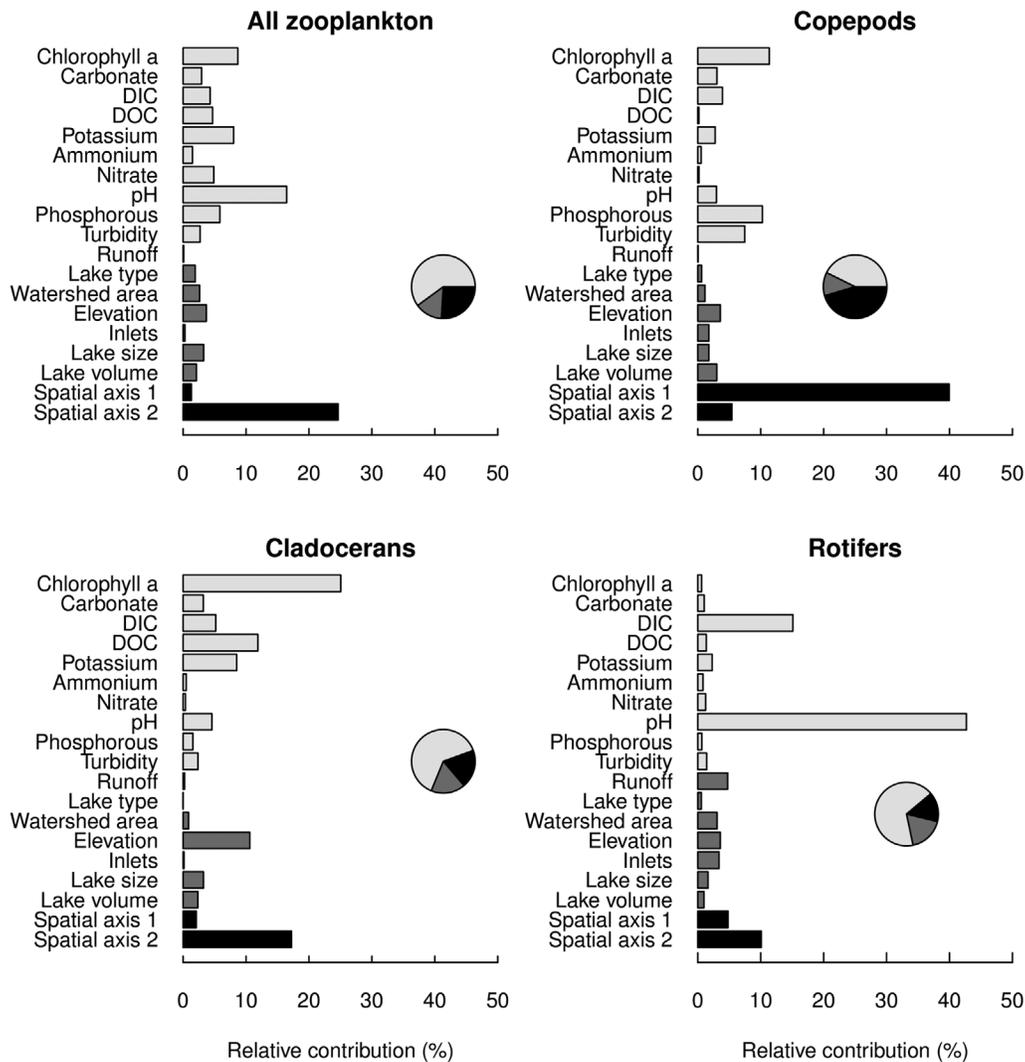


Fig. 4. The relative contributions, obtained through permutation tests of the BRT models, of geographical and environmental variables structuring the zooplankton metacommunity and subcommunities (cladocerans, copepods, and rotifers). Variables are divided into environmental (light grey), geographic (dark grey), and spatial (black). Pie charts show the relative contributions of variables to metacommunity structure.

distance among sites is not a direct measure of dispersal, and therefore do not make further claims about community assembly. However, our use of spatial distance coupled with geographic variables that are likely to impact dispersal and colonization likelihood suggests that environmental variables may be more important to the distribution and composition of the entire zooplankton metacommunity, rotifer and cladoceran subsets, but not the copepod metacommunity. It is possible that copepods are more dispersal-

limited than environmentally-limited, whereas cladocerans are limited strongly by resource availability and rotifers by pH. Taken together, our analyses suggest that zooplankton metacommunities are structured largely by environmental variables, and that species respond to environmental variables in clumped groups, where species may have similar physiological tolerances to environmental variables, or may be endemic to one geographic region. Leibold et al. (2010) found environmental factors were most impor-

tant in structuring daphniids, a finding mirrored in our analysis of the cladoceran metacommunity. However, the focus of Leibold et al. (2010) was the incorporation of phylogeny and historical biogeography on species distributions. A fruitful midpoint between the two approaches would be to incorporate phylogenetic information into the boosted regression analysis, either at the community level, as we did here, or at the individual species level, as Leibold et al. (2010) did. It is worth noting that Leibold et al. (2010) claimed that species distributions could be conflicting, such that species would respond individually across the gradient obtained in the EMS analysis. This idea is supported in our analyses for the copepod metacommunity.

The EMS framework has recently received some criticism, following an evaluation of pattern detection via null model analyses (Ulrich and Gotelli 2013). However, this criticism evaluates each of the metrics independently of one another, rather than in concert. Further, despite their critique, Ulrich and Gotelli (2013) endorse the use of coherence (referred to as “EmAbs” in their publication) and Morisita’s index, citing that coherence exhibited good type I error and that Morisita’s index exhibited good power at detecting compartmentalized structures, provided that an appropriate null model was used. Thus, while it is possible that pattern detection in metacommunities is not best assessed using the EMS framework, it currently represents one of the only mechanistic approaches, and one of the few approaches to use more than a single summary statistic. Further, this Ulrich and Gotelli (2013) critique would only influence the findings of the EMS analysis, and would not discredit the BRT analysis.

The BRT analysis suggests that variables responsible for structuring zooplankton taxonomic subsets are unique to the subset. For instance, chlorophyll *a* was associated with the cladoceran species ranges, but was relatively unimportant to the rotifer metacommunity. This suggests that cladoceran ranges may be structured by a resource gradient, as chlorophyll *a* can be used as a proxy for total algal biomass, and cladocerans feed predominantly on algae (T Taipale et al. 2008), whereas copepods are typically omnivorous (Adrian and Frost 1993, Kulkarni et al. 2013) and feed mainly on smaller particles,

including bacteria (DeMott 1982). This provides a biological rationale for chlorophyll *a* structuring cladoceran and, to a lesser extent, copepod assemblages. Further, rotifers typically feed on bacteria, and may serve as food items to some copepods and cladocerans (Williamson 1987). Aside from resource limitation, our analyses also provide evidence that pH can drive species distributions, as it has the highest relative contribution value for both the entire zooplankton metacommunity and for the rotifer metacommunity. The dominance of pH suggests that zooplankton differ in their environmental tolerance ranges for pH, a finding noted elsewhere (Holt et al. 2003), which has implications for community composition under the threat of lake acidification. Further, the impact of pH may also include the impact of other variables, as pH is often correlated with other water quality metrics such as calcium (Ca), dissolved organic carbon (DOC), and total organic phosphorus (Jeziorski et al. 2012). The importance of DOC and pH to zooplankton communities is supported by previous analyses in boreal shield lakes influenced by acidification (Derry et al. 2009).

One qualification of this study concerns geographic scale. Given the broad geographic study area, environmental and spatial variables are most likely linked through processes of spatial structuring of geographic and environmental lake variables (i.e., spatial autocorrelation). Proximity to industrial centers, which have a defined spatial structure in this system, may contribute to the degree of nitrate deposition, pH, and other water chemistry variables (Fenn et al. 2003). However, there are few ways to disentangle these coupled factors, and the BRT analysis is well suited for parsing collinear predictors. The application of BRT analysis to multivariate community data offers a way to deal with the inherent messiness of community or ecosystem scale data.

Determining the factors that shape species distributions remains a core goal of community ecology and biogeography (Holyoak et al. 2005). The current study adds to a growing body of literature concerning the mechanisms structuring species distributions (Frisch et al. 2012, Heino et al. 2012, Peres-Neto et al. 2012). Specifically, this study offers a novel way to analyze community data to address the relative impact of variables

belonging to different classes. Assessment of the relative impact of individual variables is important, as is accounting for collinearity in predictor variables, as previous studies may be confounded by the effects of both of these issues.

ACKNOWLEDGMENTS

We are grateful to A. Kramer, S. Bowden, B. McDowell, and several anonymous reviewers for comments on earlier versions of this manuscript. We also acknowledge the US EPA Environmental Monitoring and Assessment Program for their efforts collecting and openly distributing the data used in the current study. T. Dallas was supported by the Odum School of Ecology.

LITERATURE CITED

- Adrian, R., and T. M. Frost. 1993. Omnivory in cyclopoid copepods: comparisons of algae and invertebrates as food for three, differently sized species. *Journal of Plankton Research* 15:643–658.
- Anderson, M. J. 2001. Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences* 58:629–636.
- Anderson, M. J., et al. 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters* 14:19–28.
- Beisner, B. E., P. R. Peres-Neto, E. S. Lindström, A. Barnett, and M. L. Longhi. 2006. The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* 87:2985–2991.
- Clements, F. E. 1916. Plant succession. An analysis of the development of vegetation. Number 242. Carnegie Institution of Washington, Washington, D.C., USA.
- Cornell, H. V., and J. H. Lawton. 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology* 61:1–12.
- Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* 8:1175–1182.
- Cottenie, K., and L. De Meester. 2005. Local interactions and local dispersal in a zooplankton metacommunity. Pages 189–211 in M. Holyoak, M. A. Leibold, and R. D. Holt, editors. *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press, Chicago, Illinois, USA.
- Cottenie, K., E. Michels, N. Nuytten, and L. De Meester. 2003. Zooplankton metacommunity structure: regional vs. local processes in highly interconnected ponds. *Ecology* 84:991–1000.
- Dallas, T. 2013. *metacom*: Analysis of the ‘Elements of Metacommunity Structure.’ R package version 1.0. <http://CRAN.R-project.org/package=metacom>
- Dallas, T. 2014. *metacom*: an R package for the analysis of metacommunity structure. *Ecography* 37:402–405.
- De’Ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88:243–251.
- De Meester, L. 2011. A metacommunity perspective on the phylo- and biogeography of small organisms. Pages 324–334 in L. De Meester, editor. *Biogeography of microscopic organisms: Is everything small everywhere?* Cambridge University Press, Cambridge, UK.
- DeMott, W. R. 1982. Feeding selectivities and relative ingestion rates of *Daphnia* and *Bosmina*. *Limnology and Oceanography* 27:518–527.
- Derry, A. M., S. E. Arnott, J. A. Sheard, P. D. Hebert, and P. T. Boag. 2009. Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. *Ecology* 90:2275–2286.
- Elith, Jane, J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.
- Fenn, M. E. 2003. Ecological effects of nitrogen deposition in the western United States. *BioScience* 53:404–420.
- Fontaneto, D., G. F. Ficetola, R. Ambrosini, and C. Ricci. 2006. Patterns of diversity in microscopic animals: Are they comparable to those in protists or in larger animals? *Global Ecology and Biogeography* 15:153–162.
- Frisch, D., K. Cottenie, A. Badosa, and A. J. Green. 2012. Strong spatial influence on colonization rates in a pioneer zooplankton metacommunity. *PLoS ONE* 7:e40205.
- Gauch, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge, UK.
- Gilbert, B., and J. R. Bennett. 2010. Partitioning variation in ecological communities: Do the numbers add up? *Journal of Applied Ecology* 47:1071–1082.
- Gleason, H. A. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53:7–26.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.
- Gotelli, N. J., and G. R. Graves. 1996. *Null models in ecology*. Smithsonian Institution Press, Washington, D.C., USA.
- Hairton, N. G., F. E. Smith, and L. B. Slobodkin. 1960. Community structure, population control, and competition. *American Naturalist* 94:421–425.
- Havel, J. E., and J. B. Shurin. 2004. Mechanisms, effects, and scales of dispersal in freshwater zooplankton.

- Limnology and Oceanography 49:1229–1238.
- Heino, J., M. Grönroos, J. Soininen, R. Virtanen, and T. Muotka. 2012. Context dependency and metacommunity structuring in boreal headwater streams. *Oikos* 121:537–544.
- Herlihy, A. T., D. H. Landers, R. F. Cusimano, W. S. Overton, and P. J. Wigington. 1991. Temporal variability in lakewater chemistry in the northeastern United States: Results of Phase 2 of the Eastern Lake Survey. Number PB-91-159590/XAB. NSI Technology Services, Corvallis, Oregon, USA.
- Holt, C. A., N. D. Yan, and K. M. Somers. 2003. pH 6 as the threshold to use in critical load modeling for zooplankton community change with acidification in lakes of south-central Ontario: accounting for morphometry and geography. *Canadian Journal of Fisheries and Aquatic Science* 60:151–158.
- Holyoak, M., M. A. Leibold, and R. D. Holt. 2005. *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press, Chicago, Illinois, USA.
- Jeziorski, A., A. M. Paterson, and J. P. Smol. 2012. Crustacean zooplankton sedimentary remains from calcium-poor lakes: complex responses to threshold concentrations. *Aquatic Sciences: Research Across Boundaries* 74:121–131.
- Keith, S. A., A. C. Newton, M. D. Morecroft, D. J. Golicher, and J. M. Bullock. 2011. Plant metacommunity structure remains unchanged during biodiversity loss in English woodlands. *Oikos* 120:302–310.
- Kulkarni, D., A. Gergs, U. Hommen, H. T. Ratte, and T. G. Preuss. 2013. A plea for the use of copepods in freshwater ecotoxicology. *Environmental Science and Pollution Research* 20:75–85.
- Leibold, M. A., E. P. Economo, and P. Peres-Neto. 2010. Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology Letters* 13:1290–1299.
- Leibold, M. A., and G. M. Mikkelsen. 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos* 97:237–250.
- Leibold, M. A. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7:601–613.
- Medley, K. A., and J. E. Havel. 2007. Hydrology and local environmental factors influencing zooplankton communities in floodplain ponds. *Wetlands* 27:864–872.
- Mitchell-Hall, T. E., A. C. Neale, S. G. Paulsen, and J. E. Pollard. 1989. Eastern Lake Survey, Phase II: quality assurance report. EPA/600/4-85-017. U.S. Environmental Protection Agency, Las Vegas, Nevada, USA.
- Morisita, M. 1971. Composition of the *I*-index. *Researches on Population Ecology* 13:1–27.
- Mouillot, D. 2007. Niche-assembly vs. dispersal-assembly rules in coastal fish metacommunities: implications for management of biodiversity in brackish lagoons. *Journal of Applied Ecology* 44:760–767.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. R. H. Stevens, and H. Wagner. 2012. *vegan: Community Ecology Package*. R package version 2.0-4. <http://CRAN.R-project.org/package=vegan>
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Peres-Neto, P. R., M. A. Leibold, and S. Dray. 2012. Assessing the effects of spatial contingency and environmental filtering on metacommunity phylogenetics. *Ecology* 93:14–30.
- Pinel-Alloul, B., T. Niyonsenga, and P. Legendre. 1995. Spatial and environmental components of freshwater zooplankton structure. *Ecoscience* 2:1–19.
- Presley, S. J., C. L. Higgins, C. López-González, and R. D. Stevens. 2009. Elements of metacommunity structure of Paraguayan bats: multiple gradients require analysis of multiple ordination axes. *Oecologia* 160:781–793.
- Presley, S. J., C. L. Higgins, and M. R. Willig. 2010. A comprehensive framework for the evaluation of metacommunity structure. *Oikos* 119:908–917.
- Presley, S. J., and M. R. Willig. 2009. Bat metacommunity structure on Caribbean islands and the role of endemics. *Global Ecology and Biogeography* 19:185–199.
- R Development Core Team. 2013. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ridgeway, G. 2012. *gbm: generalized boosted regression models*. R package version 1.6-3.2. <http://CRAN.R-project.org/package=gbm>
- Shurin, J. B., K. Cottenie, and H. Hillebrand. 2009. Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia* 159:151–159.
- Soininen, J., M. Kokocinski, S. Estlander, J. Kotanen, and J. Heino. 2007. Neutrality, niches, and determinants of plankton metacommunity structure across boreal wetland ponds. *Ecoscience* 14:146–154.
- Taipale, S., P. Kankaala, M. Tirola, and R. I. Jones. 2008. Whole-lake dissolved inorganic ¹³C additions reveal seasonal shifts in zooplankton diet. *Ecology* 89:463–474.
- Tessier, A. J., and R. J. Horwitz. 1991. Analysis and interpretation of zooplankton species collected during Phase II of the Eastern Lake Survey. EPA-600/R-92/012. U.S. Environmental Protection Agency, Washington, D.C., USA.
- Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. *Oikos* 122:2–18.

- US EPA. 1987. Handbook of methods for acid deposition studies: laboratory analysis for surface water chemistry. EPA/600/4~7/026. U.S. Environmental Protection Agency, Washington, D.C., USA.
- US EPA. 1990. National surface water survey: Eastern Lake Survey Phase II Northeastern Lakes. EPA/60-0/3-91/028. U.S. Environmental Protection Agency, Washington, D.C., USA.
- Weiher, E., D. Freund, T. Bunton, A. Stefanski, T. Lee, and S. Bentivenga. 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B* 366:2403–2413.
- Williamson, C. E. 1987. Predator-prey interactions between omnivorous diaptomid copepods and rotifers: The role of prey morphology and behavior. *Limnology and Oceanography* 32:167–177.
- Wright, D. H., B. D. Patterson, G. M. Mikkelsen, A. Cutler, and W. Atmar. 1997. A comparative analysis of nested subset patterns of species composition. *Oecologia* 113:1–20.

SUPPLEMENTAL MATERIAL

DATA AVAILABILITY

Eastern Lake Survey data are available from the US EPA: <http://www.epa.gov/emap2/html/data/surfwatr/data/els.html>