Science of the Total Environment 609 (2017) 263–271

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Improving assessment accuracy for lake biological condition by classifying lakes with diatom typology, varying metrics and modeling multimetric indices

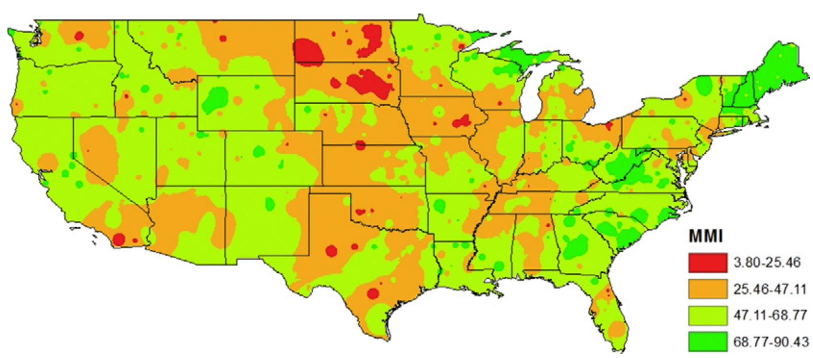
Bo Liu [a], R. Jan Stevenson [a,b,*]

[a] Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA
[b] Center for Water Sciences, Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA

HIGHLIGHTS

- Hierarchical modeling improved multimetric indices (MMI) performance.
- Modeled MMI performances were different when evaluated at different spatial scales.
- Varying metrics among site groups did not improve MMI performance.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 May 2017
Received in revised form 16 July 2017
Accepted 17 July 2017
Available online 24 July 2017

Editor: D. Barcelo

Keywords:
Diatom
Multimetric index
Lakes
Random forest
Classification and regression trees
Multiple linear regression

ABSTRACT

Site grouping by regions or typologies, site-specific modeling and varying metrics among site groups are four approaches that account for natural variation, which can be a major source of error in ecological assessments. Using a data set from the 2007 National Lakes Assessment project of the USEPA, we compared performances of multimetric indices (MMI) of biological condition that were developed: (1) with different lake grouping methods, ecoregions or diatom typologies; (2) by varying or not varying metrics among site groups; and (3) with different statistical techniques for modeling diatom metric values expected for minimally disturbed condition for each lake. Hierarchical modeling of MMIs, i.e. grouping sites by ecoregions or typologies and then modeling natural variability in metrics among lakes within groups, substantially improved MMI performance compared to using either ecoregions or site-specific modeling alone. Compared with MMIs based on ecoregion site groups, MMI precision and sensitivity to human disturbance were better when sites were grouped by diatom typologies and assessing performance nationwide. However, when MMI performance was evaluated at site group levels, as some government agencies often do, there was little difference in MMI performance between the two site grouping methods. Low numbers of reference and highly impacted sites in some typology groups likely limited MMI performance at the group level of analysis. Varying metrics among site groups did not improve MMI performance. Random forest models for site-specific expected metric values performed better than classification and regression tree and multiple linear regression, except when numbers of reference sites were small in site groups. Then classification and regression tree models were most precise. Based on our results, we recommend hierarchical modeling in future large scale

* Corresponding author.
E-mail addresses: liubo3@msu.edu (B. Liu), rjstev@msu.edu (R.J. Stevenson).

http://dx.doi.org/10.1016/j.scitotenv.2017.07.152

lake assessments where lakes are grouped by ecoregions or diatom typologies and site-specific metric models are used to establish expected metric values.

1. Introduction

Assessments of biological condition are important for managing freshwater resources (European Union, 2000; USEPA, 2007a, 2007b). In lakes, diatoms have a long history of use in paleoecological studies that document lake responses to a wide variety of human disturbances, because diatoms are sensitive to many environmental changes and current as well as past assemblages are preserved in lake sediments (Smol and Stoermer, 2010). Diatoms are also important primary producers, elements of food webs, and sources of biodiversity in lakes (Mann and Droop, 1996); thus diatoms are important elements of biological condition in lakes (sensu Davies and Jackson, 2006). Diatom assemblages may play a unique role for understanding biological integrity, because they likely respond to different types of disturbances compared to lake invertebrates and fish, as they do in streams (O'Connor et al., 2000; Hering et al., 2006; Carlisle et al., 2008; Beck and Hatch, 2009). As a result, diatoms should be particularly valuable in the assessment of current lake conditions as well as paleoecological studies.

Relationships among natural environment factors, human disturbance and metrics are complicated. Relationships between human disturbance and metrics can be influenced by the effects of natural environment on both metrics and disturbance (Stoddard et al., 2008; Hawkins et al., 2010; Schoolmaster et al., 2013). Thus, one of the challenges with assessing ecological condition across large spatial scales is distinguishing effects of human disturbance from natural variation (Stevenson et al., 2013). Natural variability in diatom assemblage composition is great at continental spatial scales and may be related to species biogeographies and the high sensitivity of diatoms species composition to naturally varying environmental factors. Stevenson et al. (2009) showed that a diatom metric for trophic status was affected as much by natural variability among streams as human disturbance. A priori classification of sites by regions or typologies, site-specific modeling of expected reference condition, and varying metrics in site groups are four approaches that have been used to control natural variation in ecological assessments (Whittier et al., 2007; Hawkins et al., 2010).

Landscape regionalizations and aquatic biota assemblage composition have been used to group sites into classes to account for natural variation among sites (Hawkins et al., 2010). Regionalization scheme, such as Omernik's ecoregions (Omernik, 1987), has been extensively used in freshwater assessment, particularly in the US (USEPA, 2010). Ecoregions and EDUs are assumed to capture a significant amount of the natural variation in metrics or multimetric indices (MMIs) caused by differences in climate, geology, hydrology, soils, and surrounding vegetation. Regionalization schemes, however, cannot account for biotic response to natural variation within an ecoregion (Hawkins et al., 2010). Biological typologies assign sites to groups (i.e. typologies) by similarity in species composition of assemblages at reference sites. Biological typologies are not spatially constrained, so they can account for natural variation within and across regions. Biological typologies are used to account for natural variation in species composition among habitats in RIVPACS (Wright et al., 2000), a widely used approach for stream bioassessment in Europe and Australia.

Site-specific modeling of expected reference condition enables adjusting individual metrics for natural variation among sites. The adjusted metric values are the difference between the unadjusted metric values and the modeled expected reference value of each metric for that site. Models for expected reference metric value for a site are calculated using reference site data including unadjusted metric values and a suite of environmental variables that are affected relatively little by humans. Up to now, a variety of statistical techniques have been used to model relationships between individual metrics or MMIs and natural gradients, such as multiple linear regression (MLR) (Stevenson et al., 2013), classification and regression trees (CART, Cao et al., 2007), and random forest (RF, Hollister et al., 2016). Linear regression and CART have advantages over other techniques, because they are easier to understand by stakeholders. But more advanced modeling techniques that involve machine learning may perform better. For example, both RF and CART can model nonlinear relationships with interactions better than MLR. Moreover, RF is less susceptible to overfitting than CART and would therefore provide more accurate predictions when used with new data than CART (Breiman, 2001; Cutler et al., 2007). The choice of technique might depend on sample size and non-linear interaction of multiple variables (Smith et al., 2013), because machine learning statistical techniques usually require larger sample sizes for precise models.

Performance of MMIs could be increased if different metrics are used in different ecoregions, because human activities and the stressors they produce vary greatly among ecoregions (Ellis and Ramankutty, 2008) and sensitivity of metrics differs among stressors (Whittier et al., 2007). Both the types and intensity of human activities vary among ecoregions, with extensive agriculture in some ecoregions and more patchy urban and agricultural activities in others (USEPA, 2013). Responses of stream diatom metrics to a nutrient dominated agricultural gradient likely differ compared to a multistressor gradient with both urban and agricultural activities (Tang et al., 2016). Whittier et al. (2007) found that using different metrics in different ecoregions provided the best MMI performance, which indicates that some metrics did not respond to human disturbance as much in some ecoregions as others. Performance of MMIs could also be increased if different metrics were used in different groups of sites defined by biological typology. For example, fish and invertebrate species richness differed in cold and warm water habitats (Mebane et al., 2003; Hughes et al., 2004). However, a trade-off exists between consistency and sensitivity when deciding whether to use different biotic metrics for MMIs in different ecoregions. MMIs might become more sensitive to human disturbance if different metrics are used among ecoregions (or site groups defined by biological typology), but changing metrics also changes what we are assessing and therefore reduces consistency in assessments across groups.

In the present study, we evaluated different methods for improving the performance of a nationwide diatom MMI for lakes with the US Environmental Protection Agency's dataset from the 2007 National Lakes Assessment (NLA). We evaluated three hypotheses: (1) performance of MMIs will be greater when grouping sites by diatom typology than by ecoregions; (2) MMIs generated by selecting metrics for each site group (typologies or ecoregions) will perform better than by using the same set of metrics in all site groups; and (3) different statistical techniques (e.g. MLR, CART, RF) for adjusting metrics for natural variability will perform best in different situations. To do this, we grouped sites by ecoregions and diatom typology and calculated site-specific models of expected reference condition for each group of sites by ecoregion or typology. We then compared metric and MMI performance using a standard set of statistics that have been used in other evaluations of ecological assessment methods.

2. Materials and methods

2.1. Data sets

The NLA was conducted by the United States Environment Protection Agency (USEPA). The NLA provides a nationwide dataset and