Science of the Total Environment 645 (2018) 895–900

Contents lists available at ScienceDirect

Science of the Total Environment

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

Advancing evaluation of bioassessment methods: A reply to Liu and Cao

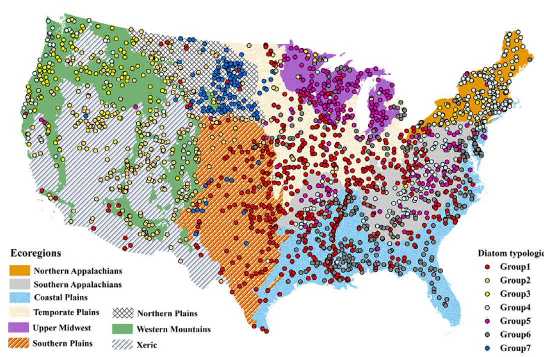
R. Jan Stevenson [a,*], Tao Tang [b], Dana M. Infante [c]

[a] Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA
[b] Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China
[c] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA

HIGHLIGHTS

- Diatom multimetric indices were evaluated with large-scale transcontinental surveys.
- Reliability of assessment improvements were evaluated with independent datasets.
- Site-specific metric modeling (SSMM) greatly improved MMI performance.
- With SSMM, MMIs by ecoregions performed as well as by typologies.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 9 July 2018
Accepted 10 July 2018
Available online xxxx

Editor: D. Barcelo

Keywords:
Diatoms
Multimetric indices
Bioassessment
Assessment

ABSTRACT

A series of three papers was written about the development of multimetric indices (MMIs) using diatoms in rivers, streams and lakes for transcontinental surveys conducted by the United States Environmental Protection Agency. Stevenson et al. (2013) used the surface sediment diatom data from the 2007 National Lake Assessment to develop national scale site specific models for MMIs to account for natural variation in condition among sites. Liu and Stevenson (2017) also used the 2007 lakes data to evaluate performance of MMIs by grouping sites by ecoregions or typologies (naturally similar types of lakes defined by similarity in diatom species composition) with site specific metric models (SSMMs) that adjust metrics for natural variability among sites. Tang et al. (2016) used benthic diatom data from the 2008–2009 National River and Stream Assessment to develop SSMMs and MMIs by ecoregion and typology. All three studies showed that SSMMs improved performance of diatom MMIs by accounting for natural variation among sites. None of the studies provided consistent evidence that grouping sites by typologies produced better MMI performance than grouping sites by ecoregions.
Liu and Cao (2018) criticized the Tang et al. (2016) paper for using means and standard errors to evaluate relative performance of MMI calculation methods at the site group scale, however, their criticism is incorrect. Actually, Tang et al. (2016) only used means to summarize and report relative performance of MMI calculation methods in the body of the paper. Tang et al. (2016) appropriately used non-parametric rank sum approaches to evaluate the probability that the multiple MMI calculations for separate site groups were the same for ecoregion (n = 9) and typology (n = 7) site groups. Liu and Stevenson (2017) used this same non-parametric approach for tests of lake diatom MMIs. Liu and Cao's (2018) concerns can be addressed by distinguishing between the goals and methods used for testing and evaluation of MMI calculation methods at the national and site-group scales.

* Corresponding author.
E-mail address: rjstev@msu.edu (R.J. Stevenson).

https://doi.org/10.1016/j.scitotenv.2018.07.129

Tang et al. (2016) did not aggregate data across site groups to test MMI performance at the national scale because they were following standard EPA methods that develop separate MMIs for each site group. In conclusion, Liu and Cao (2018) misunderstood the MMI evaluation in Tang et al. (2016) and added no new information to this body of work, because all the concerns they raised were discussed in Liu and Stevenson (2017).

1. Introduction

Tang et al. (2016) evaluated whether different methods for accounting for natural variation among sites improved performance of multimetric indices (MMIs) in a transcontinental survey of rivers and streams of the United States. Tang et al. (2016) evaluated repeatability in performance differences for five assessment methods with independent characterizations of MMI performance for multiple groups of sites. Liu and Cao (2018) argued that results of Tang et al. (2016) "are based on a statistically flawed method and conclusions drawn by them are questionable."

In this reply, we explain how Liu and Cao misunderstood the scale of analysis, hypotheses, and rationale for the analytical methods in Tang et al. (2016). Our reply will first provide a little background about assessment methods commonly used in the United States so that the terminology in our reply is clear. We will then describe some advances in ecological assessment methods used in the National Aquatic Resource Surveys (NARS) by the United States Environmental Protection Agency (US EPA) over the past decade and three related papers on diatom assessment methods involving Stevenson and this US EPA program. With this background established, our reply will show analyses by Tang et al. (2016) are not flawed because Liu and Cao (2018) misunderstood the scale of analysis and the related objective for evaluating performance of multimetric indices (MMIs) at that scale. This same analysis was repeated in the paper on lake diatom MMIs by Liu and Stevenson (2017). Our reply will show that all the issues raised by Liu and Cao (2018) are already discussed in Liu and Stevenson (2017), but Liu and Cao (2018) left out the many key points discussed in Liu and Stevenson (2017) that reconcile the differences in results between Tang et al. (2016) and Liu and Stevenson (2017). Thus the comment from Liu and Cao (2018) provided no new information.

Our reply is relatively long and thorough, because Liu and Cao (2018) sufficiently confused hypotheses tested and methods used for evaluating hypotheses that reviewers and therefore other readers may also be confused about the contributions that were made in the papers by Tang et al. (2016) and Liu and Stevenson (2017). Both papers were published in Science of the Total Environment. Our reply also provides a short synthesis of methods for evaluating assessment measurement performance and results from the three diatom MMI papers using US EPA NARS data.

2. Background

2.1. Multimetric index development and improvements in the US EPA NARS

Multimetric indices of biological condition have a long history of use in the United States to provide a broad characterization of the structure and function of biological assemblages in aquatic habitats (Karr, 1981; Barbour et al., 1999; Stoddard et al., 2008). Different metrics quantitatively characterize different elements of biological assemblages, and then the metrics are summarized in a single multimetric index. Development of multimetric indices (MMIs) involves selecting attributes of biological assemblages that differ most between sites with minimal and high human disturbance. Selected attributes, called metrics (sensu Karr and Chu, 1999), should also measure multiple aspects of biological assemblages that represent different elements of biological condition (sensu Davies and Jackson, 2006), and they should not be highly correlated. We are focusing on methods for assessing aquatic ecosystems in the United States, because the diatom MMI papers related to Liu and Cao's (2018) comment were designed to advance assessment methods for the US EPA NARS.

Advances in assessment of aquatic ecosystems have addressed issues of sensitivity, accuracy, and consistency. Here sensitivity is magnitude of assessment response to human disturbance; accuracy characterizes how assessments measure effects of human disturbance without interference by natural variability among sites; and consistency characterizes how assessments measure the same attributes or use the same methods among groups of sites and among assessments. Sensitivity of assessments has been improved by selecting a subset of attributes of ecological condition that have high separation power between sites with minimal and high levels of human disturbance (Stoddard et al., 2008). Sensitivity of assessments can also be improved by accounting for natural variation among sites and improving precision (reducing variability) in metrics and MMI values among reference sites. Thus, separation power and precision are two measures used to compare MMI performance in Stevenson et al. (2013), Tang et al. (2016), and Liu and Stevenson (2017). Accuracy of assessments to measure human disturbance should be improved by accounting for effects of natural variability on minimally disturbed ecological condition and on how natural variation in ecological systems affects their response to human disturbance (Schoolmaster et al., 2013). Consistency in assessments has been improved by using the same analytical methods in all subgroups of sites in large scale assessments (Stoddard et al., 2008; Whittier et al., 2007). Improvements in consistency in assessments have also been explored by using the same metrics for all sites for large-scale assessments and by using one model for detecting deviation from minimally disturbed or expected condition (Stevenson et al., 2013).

Ecoregions have been used to account for natural variation among habitats and improve sensitivity and accuracy of many assessments of the US (Omernik, 1987; Hughes and Larsen, 1988; Barbour et al., 1999; Crawford et al., 2016). In the US EPA's Wadeable Streams Assessment (US EPA, 2006), biological condition was assessed with MMIs based on benthic macroinvertebrates and by using ecoregions to account for natural variation among sites (Stoddard et al., 2008). Multimetric indices for the Wadeable Streams Assessment (WSA) were developed and tested separately for each ecoregion, and MMIs often used different metrics in different ecoregions (Table 1). The WSA also used the Macroinvertebrate Observed/Expected (O/E) Ratio of Taxa Loss, which provides a consistent assessment of one important attribute of ecological condition; and it also accounts for natural variability among sites (Hawkins et al., 2000).

Typologies are another method for grouping sites to account for natural variation in ecological assessments (Table 1). They are important for calculating the proportion of reference taxa that are expected (E) to occur at a site in comparison to those that were observed (O) at a site (i.e., O/E; Hawkins et al., 2000). In this case, the number of reference taxa at a site is first determined by a typological classification of reference sites, which calls for clustering reference sites with similar species composition and determining the natural factors that are important determinants for assigning sites to these clusters (i.e. typologies of sites). Tang et al. (2016) argued that typologies should better identify the natural factors regulating a group of aquatic organisms than ecoregions. Ecoregions are a form of typology, but they are spatially constrained to define a region, and they are based on natural