Quantitative Identification of Disturbance Thresholds in Support of Aquatic Resource Management

Travis O. Brenden · Lizhu Wang · Zhenming Su

Abstract The identification of disturbance thresholds is important for many aspects of aquatic resource management, including the establishment of regulatory criteria and the identification of stream reference conditions. A number of quantitative or model-based approaches can be used to identify disturbance thresholds, including nonparametric deviance reduction (NDR), piecewise regression (PR), Bayesian changepoint (BCP), quantile piecewise constant (QPC), and quantile piecewise linear (QPL) approaches. These methods differ in their assumptions regarding the nature of the disturbance-response variable relationship, which can make selecting among the approaches difficult for those unfamiliar with the methods. We first provide an overview of each of the aforementioned approaches for identifying disturbance thresholds, including the types of data for which the approaches are intended. We then compare threshold estimates from each of these approaches to evaluate their robustness using both simulated and empirical datasets. We found that most of the approaches were accurate in estimating thresholds for datasets with drastic changes in responses variable at the disturbance threshold. Conversely, only the PR and QPL approaches performed well for datasets with conditional mean or upper boundary changes in response variables at the disturbance threshold. The most robust threshold identification approach appeared to be the QPL approach; this method provided relatively accurate threshold estimates for most of the evaluated datasets. Because accuracy of disturbance threshold estimates can be affected by a number of factors, we recommend that several steps be followed when attempting to identify disturbance thresholds. These steps include plotting and visually inspecting the disturbance-response data, hypothesizing what mechanisms likely generate the observed pattern in the disturbance-response data, and plotting the estimated threshold in relation to the disturbance-response data to ensure the appropriateness of the threshold estimate.

Keywords Threshold · Land use · Nutrient · Stream · Fish indicator

Introduction

Anthropogenic disturbances, such as urban development within watersheds and industrial or municipal wastewater generation, can lead to increased nutrient, sediment, and toxicant inputs to aquatic ecosystems and can modify the physicochemical, thermal, hydrologic, and morphologic characteristics of streams (Paul and Meyer 2001; Wang and Lyons 2003). Such disturbances also can negatively affect the structure and function of biological communities (Karr and Chu 1999; Wang and others 2003). Understanding how aquatic communities respond to increasing levels of disturbance is critical for many aspects of stream management, including assessing stream health, predicting future risks, rehabilitating degraded waterbodies, and establishing regulatory criteria.

When undisturbed, stream physicochemical, morphological, thermal, and hydrological conditions are in a

T. O. Brenden
Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, 153 Giltner Hall, East Lansing, MI 48824, USA

L. Wang (✉) · Z. Su
Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 212 Museums Annex, Ann Arbor, MI 48109, USA
e-mail: lizuwang@umich.edu

dynamic equilibrium, with conditions varying around long-term averages in response to fluctuations in large-scale factors such as climate (Wang and others 2006a). Biological communities in streams have evolved under such conditions and consequently can be quite resilient to minor disturbances. Under intensified disturbance, however, ecosystem functions and biological communities can become increasingly stressed until some critical threshold is surpassed, whereupon drastic changes in stream condition can result from even slight changes in disturbance levels (Fore and others 1996; Allan 2004; Wang and others 2006b; Groffman and others 2006). Stream ecologists, managers, and policy makers are greatly interested in identifying disturbance thresholds for establishing regulatory criteria for aquatic systems and for identifying stream reference conditions against which stream degradation can be judged (Wang and others 2007).

Disturbance thresholds are routinely identified through visual inspection of data series plots (Schueler 1994; Booth and Jackson 1997; Wang and others 2003) or by setting thresholds equal to some percentile (e.g., 75th percentile) of the observed disturbance levels (EPA 2000a, b). There are perceived problems with both of these approaches. With the visual inspection approach, different researchers may identify different thresholds for the same dataset due to the subjective nature of the process. Identification of disturbance thresholds by the percentile approach can be problematic because of differing levels of disturbance among regions. A region with relatively few disturbed waterbodies may require a different percentile to accurately estimate the disturbance threshold than a region with lots of disturbed waterbodies.

Several quantitative or model-based approaches to identifying thresholds have been proposed, and the interest in using such methods is increasing. Proposed approaches for identifying disturbance thresholds include piecewise regression (Toms and Lesperance 2003), Bayesian change-point analysis (Qian and others 2003, 2004), and nonparametric deviance reduction assessment (Qian and others 2003). Two other methods that, as far as we are aware, have not been used to identify disturbance thresholds, but which seem promising, are quantile piecewise constant and quantile piecewise linear approaches (Chaudhuri and Loh 2002). Although each of the aforementioned methods can be used to identify disturbance thresholds, they differ in their assumptions regarding the nature of the disturbance-response variable relationships, which can affect the accuracy of the threshold estimates. To those unfamiliar with these methods, it may not be clear what assumptions are made by the approaches regarding the disturbance-response variable relationship or how results might vary if several of the approaches were applied to the same set of data.

The purpose of this research was to present an overview of each of the aforementioned approaches for identifying

disturbance thresholds, including the types of disturbance-response variables datasets for which the methods are appropriate. Additionally, using simulated and empirical datasets, we compare threshold estimates among each of these approaches as a way of evaluating the robustness of the methods. We felt that this comparison would be beneficial in selecting among the approaches as factors such as small sample sizes and the occurrence of measurement error may make it difficult to know what the underlying disturbance-response variable relationship is for a particular set of data.

Methods

Overview of Techniques

Nonparametric deviance reduction (NDR) is an approach for threshold identification that entails splitting a disturbance-response dataset into two groups at points along the ordered disturbance variable gradient and calculating the reduction in the response variable deviance that results from the split (Qian and others 2003). For a continuous variable, the deviance can be calculated as

$$D = \sum_{i=1}^{n} (y_i - \mu)^2$$

where D is the deviance, n is the sample size, and μ is the mean of the n response variable observations (Qian and others 2003). For categorical data, the deviance can be calculated as

$$D = -2 \sum_{i=1}^{g} n_i \log (p_i)$$

where g is the number of classes, p_i is the proportion of observations, and n_i is the number of observations in class i (Qian and others 2003). The split that results in the largest reduction in the deviance is the disturbance threshold estimate. According to Qian and others (2003), the NDR approach to threshold identification is inspired by tree-based modeling; the identified threshold is in fact the first split of a regression tree model. The type of disturbance-response threshold pattern assumed by the NDR approach can be best described as a stair-step relationship in which there is a drastic change in response variable characteristics at the disturbance threshold (Fig. 1A). The NDR approach was used by Qian and others (2003) to identify soil phosphorous thresholds based on differences in macorinvertebrate communities. Wang and others (2007) also used the NDR approach to identify nutrient thresholds based on differences in fish and macroinvertebrate occurrence.

The Bayesian changepoint (BCP) approach for identifying thresholds conceptually is similar to the NDR approach in that it entails splitting a disturbance-response dataset into two or more groups at points along the ordered disturbance variable gradient. Thus, the BCP approach would also work well for a stair-step disturbance-response threshold pattern. Unlike the NDR approach, however, the BCP approach requires explicit specification as the probability distributions of the groups that result from splitting the original dataset. For example, if one assumes that the response
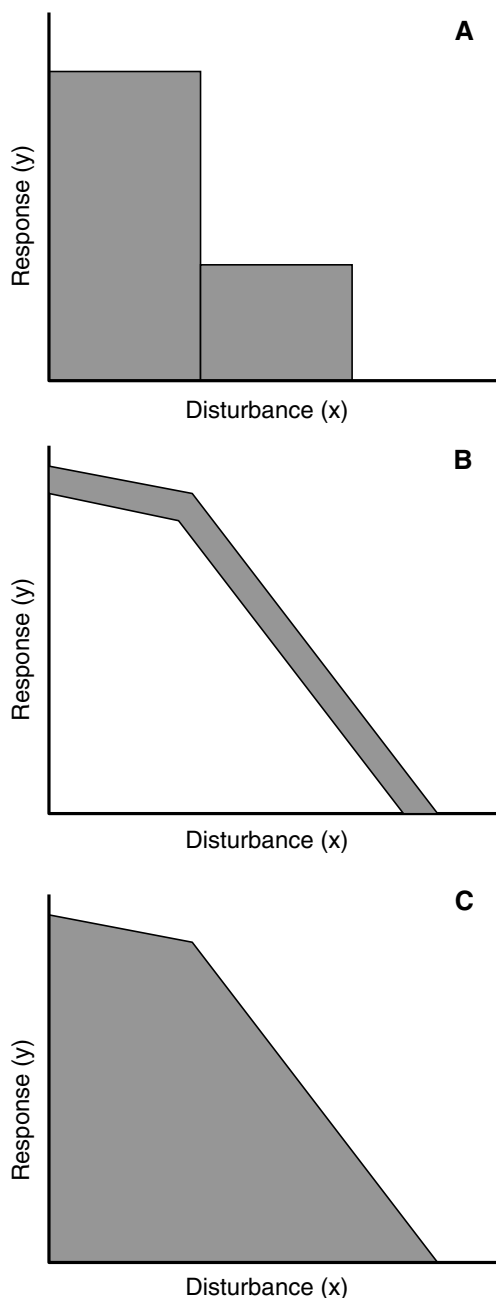

Fig. 1 Illustration of three disturbance-response threshold patterns that may be encountered in aquatic resources management. Top panel, stair step; middle panel, condition mean; bottom panel, wedge-shaped

variable is distributed normally, then splitting the dataset into two groups would result in the groups having normal distributions with means and variances of μ_1 and σ_1^2 and μ_2 and σ_2^2. Alternatively, if one assumed that the response variable has a binomial distribution, then splitting the datasets into two groups would result in the groups having binomial distributions with parameters p_1, N_1, p_2, and N_2, where p_i is the probability of success for group i ($i = 1$ or 2) and N_i is the number of sites for the groups. Because the BCP approach is based on Bayesian statistical principles, prior distributions can be assigned to all model parameters, including the threshold estimate, which can be useful when previous research has been conducted on a particular disturbance-response variable relationship. If prior research has not been conducted, then uninformative priors can be specified for the model parameters. Qian and others (2003) used Bayesian changepoint analysis for both binomial and normally distributed data to identify soil phosphorous thresholds for macroinvertebrate communities. Qian and others (2004) used Bayesian changepoint analysis for multinomial distributed data to identify soil phosphorous thresholds for diatoms and macroinvertebrates.

Piecewise regression (PR) models are "broken-stick" models that consist of two or more separate regression lines describing the conditional mean relationship between the disturbance and response variables. The regression lines are joined at unknown points along the ordered gradient of the disturbance variable (Toms and Lesperance 2003). The points where the regression lines are joined, which are referred to as knots, breakpoints, or joinpoints, indicate where there are changes in the conditional mean relationship between the disturbance and response variables and are interpreted as threshold estimates for the disturbance variable. Mathematically, piecewise regression models can be represented as

$$y_i = \begin{cases} \beta_{10} + \beta_{11}x_i, & x_i \leq \alpha \\ \beta_{20} + \beta_{21}x_i, & x_i > \alpha \end{cases}$$

where β_{10} and β_{11} are the regression intercept and slope, respectively, for the first group of observation, β_{20} and β_{21} are the regression intercept and slope for the second group of observations, and α is the threshold estimate. Continuity among the pieces is imposed through the constraint $\beta_{10} + \beta_{11}\alpha = \beta_{20} + \beta_{21}\alpha$ (Seber and Wild 2003). To fit piecewise regression models by nonlinear least squares, we want to find estimates of β_{10}, β_{11}, β_{20}, β_{21}, and α such that the quantity

$$\sum_{i=1}^{n} (y_i - \mu(x_i, \beta, \alpha))^2$$

is minimized, with $\mu(x_i, \beta, \alpha)$ representing the piecewise conditional mean function from above. Fitting PR models

can be difficult due to the occurrence of many local minima in the objective function surface, which is one of the challenges in using this method to identify disturbance thresholds. The type of disturbance-response threshold pattern assumed by the PR approach can be best described as a conditional-mean relationship in which there is a change in the conditional mean of the response variable per unit increase at the disturbance threshold (Fig. 1B). Toms and Lesperance (2003) used the PR approach to identify forestry harvest thresholds for understory plant community occurrence. Denoël and Ficetola (2007) used the PR approach to identify landscape configuration thresholds for newt conservation.

The quantile piecewise constant (QPC) is similar to the NDR approach in that it is a tree-based method for identifying disturbance thresholds. The NDR and QPC approaches differ, however, in terms of what characteristics of the response variable are used to partition the groups. With the QPC approach, trees are generated by recursively dividing the data series such that there are maximum differences in the τth quantiles of the response variables. The use of response variable quantiles to generate regression trees is not new. Breiman and others (1984) proposed constructing median regression trees as a robust alternative to mean-based regression trees. Chaudhuri and Loh (2002) extended this concept to deal with a broader range of quantiles. As with the NDR approach, the QPC approach for identifying disturbance thresholds would work well for a stair-step disturbance-response threshold pattern. One advantage of identifying thresholds with the QPC approach is that one can more fully explore the characteristics of the response variable. For example, quantile regression trees could be constructed based on the 50th, 75th, and 95th percentiles of the response variable and the rate of change in threshold estimates could be considered, which might lead to greater insight regarding whether a disturbance threshold truly exists for a particular dataset.

The quantile piecewise linear (QPL) approach for identifying disturbance thresholds blends characteristics of the QPC and PR approaches. Like the QPC approach, the QPL approach is a regression tree method that uses quantiles to partition groups. Whereas the QPC approach generated trees by recursively partitioning a data series such that each tree node has maximum differences in the τth quantiles of the response variables, the QPL approach partitions the data based on differences in the conditional relationship between the disturbance and response variables at each tree, which is somewhat similar to the PR approach for threshold identification. Unlike the PR approach, however, the QPL approach is capable of identifying thresholds based on the lower and upper tails of the conditional distribution of the response variables.

Mathematically, the QPL approach can be represented with the same equation used to represent the PR approach. Fitting of the β_{10}, β_{11}, β_{20}, β_{21}, and α parameters for the QPL approach, however, entails minimization of the sum

$$\sum_{i=1}^{n} \{|y_i - \xi(x_i, \beta, \alpha)| + (2\tau - 1)(y_i - \xi(x_i, \beta, \alpha))\}$$

where $\xi(x_i, \beta, \alpha)$ is the piecewise conditional quantile function (Chaudhuri and Loh 2002). Unlike the SR approach, the QPL approach does not impose a continuity restriction on the regression pieces; thus, in addition to there being differences in regression slopes, there also may be jumps or falls among the regression lines at the disturbance thresholds. There are several disturbance-response variable relationships for which the QPL approach would work well; one such pattern would be a wedge-shaped dataset (Fig. 1C) in which there was a change in the upper boundary relationship between the disturbance and the response variables at some threshold. Wedge-shaped datasets are believed to be common in aquatic resource management (Cade and others 1999; Wang and others 2003, 2007) and result from response variables being affected by several potentially limiting factors, only some of which are measured, with the upper limits of the response variable being set by one of the measured factors (Cade and Noon 2003).

Data Simulation and Collection

We simulated datasets for the stair-step, conditional mean, and wedge-shaped disturbance-response threshold patterns described in the previous section to evaluate the accuracy and robustness of the different threshold identification approaches. Each simulated dataset consisted of 500 observations. We simulated a stair-step disturbance-response threshold pattern by randomly drawing data from uniform probability distributions. At disturbances <0.2, the response variable could take on values between 0.0 and 1.0. At disturbances >0.2, however, the response variable could only take on values between 0.0 and 0.2. We simulated a conditional-mean dataset using the equation

$$y_i = \begin{cases} 0.9, & x_i \leq 0.2 \\ 1.20 - 1.5x_i, & x_i > 0.2 \end{cases}$$

In other words, the response variable was constant at disturbances <0.2, but at disturbances >0.2 there was a 1.5 times decrease in the conditional mean of the response per unit increase in the disturbance variables. Random errors generated from a normal probability distribution with a mean of 0 and a standard deviation of 0.1 were added to the simulated data to reflect measurement uncertainty for both the response and the disturbance variables. We simulated a

wedge-shaped dataset using an interference interaction model as described by Cade and others (1999). The wedge-shaped dataset was such that there was a change in the upper boundary relationship between the disturbance and the response variables at disturbances >0.2 (Fig. 2).

The empirical datasets that we used to compares threshold estimate among the different quantitative approaches were collected as part of previous research conducted in Michigan (Wang and others 2008) and Wisconsin (Wang and others 2007). The main objectives for the Michigan research were to identify stream reference sites and to quantify human disturbance gradients for streams within the state (Wang and others 2008). The

Michigan dataset consisted of measurements from 741 wadable streams (defined as streams with network watershed areas <1,600 km^2 or stream orders <fifth order) located throughout the state (Wang and others 2008). Disturbance attributes measured for the streams included road density (RDDENS; km/km^2), population density (POPDENS; people/km^2), and percentage urban land use (URBAN; %) in the watersheds upstream of the sites. Two biological indicators of stream health were calculated for the stream sites using fish assemblage data collected by electrofishing: percentage of captured fish that were human disturbance intolerant (INTOL; %) and an index of biotic integrity (IBI) score. IBI scores for coldwater streams were calculated using an IBI procedure developed for Wisconsin coldwater streams (Lyons and others 1996); IBI scores for warmwater streams were calculated using an IBI procedure developed specifically for Michigan warmwater streams (Wang and others 2008). For Michigan streams not considered either cold- or warmwater, both cold- and warmwater IBI scores were calculated and the higher of the two scores were assigned to the streams.

The main objectives for the Wisconsin study were to examine the relationships between nutrient concentration and fish community composition, and to evaluate the importance of nutrient concentration in structuring fish community composition relative to other physicochemical variables (Wang and others 2007). Data from the Wisconsin study were from 235 streams located throughout the state (Wang and others 2007). Disturbance attributes measured for the streams included total phosphorous (TP; mg/l), dissolved phosphorus (DP; mg/l), and dissolved ammonia nitrogen (NH$_4$; mg/l) concentrations. Three biological indicators of stream conditions were calculated from fish assemblage data collected by electrofishing: percentage of captured fish that were human disturbance intolerant (INTOL; %), number of captured fish from the Salmonidae family (SALMN), and IBI scores (Wang and others 2007). IBI scores for Wisconsin coldwater streams were calculated using the Wisconsin coldwater IBI procedure (Lyons and others 1996), while IBI scores for warmwater streams were calculated using a Wisconsin warmwater procedure (Lyons 1992). As was done for Michigan streams, both cold- and warmwater IBI scores were calculated for streams not considered either cold- or warmwater and the higher of the two scores were assigned to streams.
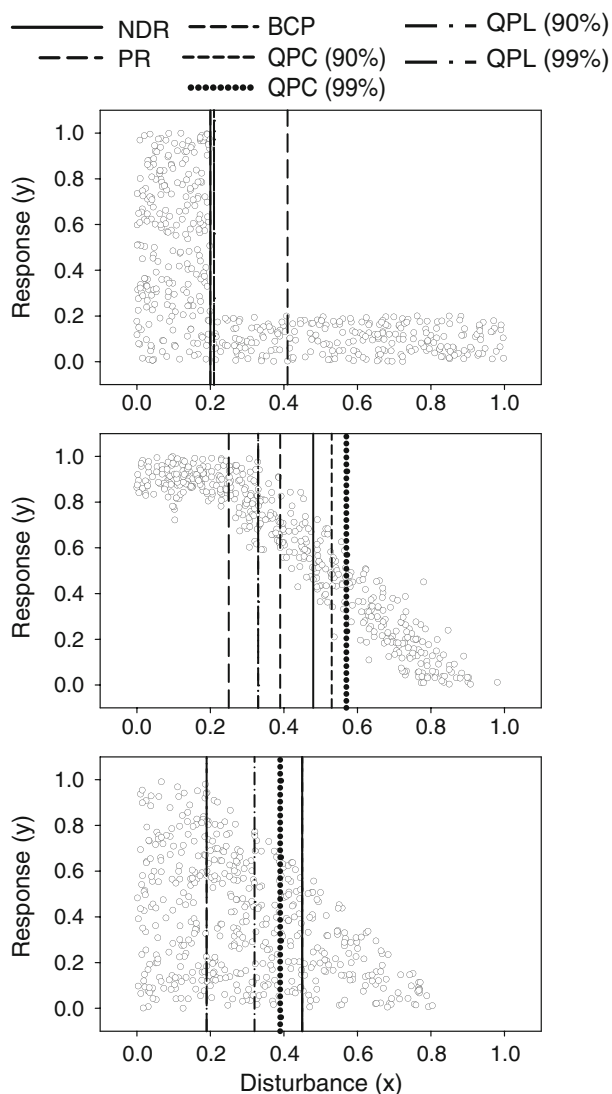

Fig. 2 Scatterplots of the simulated datasets used in evaluating the accuracy of the threshold identification methods (unitless). Locations of the threshold estimates from each of the identification methods are also shown on the plots. Top panel, simulated stair-step dataset; middle panel, simulated condition mean dataset; bottom panel, simulated wedge-shaped dataset

Data Analysis

Disturbance thresholds for the simulated and empirical datasets based on the NDR approach were calculated in R (R Development Core Team 2007) using the 'rpart' package (Therneau and others 2007), which is an R package for constructing classification and regression trees. We used

the first split identified by the 'rpart' regression tree as our NDR threshold estimate. Disturbance thresholds based on the BCP approach were estimated in WinBUGS (Lunn and others 2000), which uses Gibbs sampling to construct a Markov chain for determining the joint posterior probability distribution for model parameters. We assumed that we had no knowledge as to the location of disturbance thresholds and thus used uniform distributions as threshold priors when fitting all models. We used the median of the marginal posterior distribution of the threshold parameter as the threshold point estimate. For the simulated datasets, we assumed that the response variables were distributed normally. We also assumed that the IBI response variables for the empirical datasets were distributed normally. We assumed that the INTOL response variables from the empirical datasets were distributed binomially. We assumed a Poisson distribution for the SALMN response variable for the Wisconsin empirical dataset.

Disturbance thresholds for the PR approach were estimated in SAS (SAS Institute 2003) using the NLIN procedure. The PR models were fit by grid-search, which is a recommended approach for fitting these types of models (Lerman 1980) because of the occurrence of many local minima in the objective function surface. In some cases, it was necessary to specify a very fine sampling grid in order to keep the models from converging on local least-squares minima. Other software packages also could have been used to fit the PR models, including R through the 'segmented' package (Muggeo 2003) or the Joinpoint software program (National Cancer Institute 2005). We elected to use SAS to fit the PR models to have better control over the sampling grid during the model fitting process.

Disturbance thresholds for the QPC and QPL approaches were estimated using GUIDE (Loh 2002, 2007), which is a flexible program for fitting a wide range of classification and regression tree models, including quantile regression trees. As with the NDR approach, we assumed that the first splits identified by the GUIDE program were the QPC and QPL threshold estimates. Disturbance thresholds for the QPC and QPL approaches were estimated based on the 90th and 99th percentiles of the response variables; thus, we were primarily interested in exploring the benefits of considering upper boundary differences in response variables as a way to identify disturbance thresholds.

When estimating the disturbance thresholds for the simulated datasets and the IBI response variables for the Michigan and Wisconsin datasets by the NDR, PR, QPC, and QPL approaches, the response variables were not transformed. For the INTOL response variable, which was expressed as a percentage, the response variable was transformed using a logit transformation for the NDR, PR, QPC, and QPL approaches. For the SALM response

variable in the Wisconsin dataset, which was expressed as a count, the response variable was $\log_e$ transformed prior to analysis with the NDR, PR, QPC, and QPL approaches.

For the simulated datasets, we evaluated the performance of the threshold identification methods by comparing actual thresholds with the thresholds estimated by each of the approaches. For the empirical datasets, because we did not know whether a disturbance threshold existed for the datasets or where their exact locations were, we simply evaluated consistency of threshold estimates among the approaches as a way of gauging expected differences among the methods.

Results

Simulated Data

For the simulated stair-step dataset, the threshold estimates from the NDR, BCP, QPC, and QPL approaches were very close to the actual threshold for this dataset (Table 1). The threshold estimates for the NDR and BCP approaches equaled 0.20, while for the QPC and QPL approaches, the threshold estimates equaled 0.21. The fact that the NDR, BCP, and QPC approaches accurately estimated the threshold was not particularly surprising given that these approaches are intended for this type of disturbance-response threshold dataset. The accuracy of the QPL approach was slightly more surprising. The threshold estimate for the PR approach was 0.41 for the simulated stair-step dataset, indicating that this approach did not perform well for this type of disturbance-response threshold pattern.

For the simulated conditional-mean dataset, the PR approach yielded the most accurate threshold estimate (Fig. 2), which again was not surprising given that the PR approach was intended for this type of data. The threshold estimate from the PR approach equaled 0.25, which was only slightly greater than the actual threshold of 0.20. The QPL 90% and 99% methods gave the next most accurate threshold estimates; both methods gave threshold estimates of 0.33. Based on plots of the threshold estimates in relation to the simulated data, it appeared that the accuracy of the QPL 90% and 99% approaches was affected by the addition of random errors to the simulated data series (Fig. 1). It is possible that using a different quantile (e.g., 50% or 80%) may have improved the accuracy of this approach. The BCP, NDR, and QPC approaches gave the least accurate threshold estimates for the conditional-mean dataset; their threshold estimates ranged from 0.39 to 0.57 (Table 1).

For the simulated wedge-shaped dataset, the SR and QPL 99% method yielded the most accurate threshold estimates (Fig. 2). The threshold estimates from these approaches equaled 0.19, which was just slightly less than

Table 1 Estimated thresholds for the simulated, Michigan, and Wisconsin datasets

Variable	NDR	PR	BCP	QPC 90%	QPC 99%	QPL 90%	QPL 99%
Simulated dataset							
Stair-step	0.20	0.20	0.41	0.21	0.21	0.21	0.21
Conditional mean	0.48	0.39	0.25	0.53	0.57	0.33	0.33
Wedge-shaped	0.45	0.45	0.19	0.45	0.39	0.32	0.19
Michigan dataset							
INTOL vs. ROAD	5.58	1.0	1.10	0.77	2.04	0.69	2.05
INTOL vs. POP	12.54	12.5	16.73	12.54	11.28	13.54	10.86
INTOL vs. URB	1.86	3.1	0.82	0.51	2.33	0.52	2.29
IBI vs. ROAD	1.81	1.6	0.34	0.73	1.16	0.69	1.16
IBI vs. POP	12.92	12.9	10.12	12.54	19.57	12.80	5.13
IBI vs. URB	0.12	3.5	0.13	0.98	1.51	0.98	1.51
Wisconsin dataset							
IBI vs. TP	0.39	0.07	0.03	0.04	0.06	0.04	0.07
IBI vs. DP	0.30	0.05	0.07	0.06	0.04	0.02	0.04
IBI vs. NH_4	0.02	0.02	0.03	0.02	0.04	0.07	0.04
INTOL vs. TP	0.16	0.10	0.08	0.11	0.06	0.11	0.06
INTOL vs. DP	0.01	0.04	0.10	0.07	0.04	0.07	0.04
INTOL vs. NH_4	0.06	0.03	0.05	0.03	0.03	0.07	0.03
SALMN vs. TP	0.14	0.09	0.12	0.09	0.07	0.09	0.13
SALMN vs. DP	0.08	0.05	0.09	0.06	0.06	0.04	0.09
SALMN vs. NH_4	0.05	0.02	0.03	0.02	0.03	0.03	0.05

Note. NDR—nonparametric deviance reduction; PR—piecewise regression; BCP—Bayesian change point; QPC 90%—quantile piecewise constant (90th percentile); QPC 99%—quantile piecewise constant (99th percentile); QPL 99%—quantile piecewise linear (90th percentile); QPL 99%—quantile piecewise linear (99th percentile); INTOL—percentage of captured fish that were human disturbance intolerant (%); IBI—fish index of biotic integrity; SALMN—number of captured fish that belonged to the Salmonidae family; ROAD—watershed road density (km/km^2); POP—watershed population density (people/km^2); URBAN—watershed urban land use (%); TP—total phosphorus (mg/l); DP—dissolved phosphorous (mg/L); NH_4—dissolved ammonia nitrogen (mg/l)

the actual threshold of 0.20. The QPL 90% approach was the next most accurate method, provided a threshold estimate of 0.32. As was found with the conditional mean dataset, the BCP, NDR, and QPC approaches provided the least accurate estimates, with thresholds ranging from 0.39 to 0.48 (Table 1).

Michigan Data

The disturbance thresholds identified for the INTOL-RDDENS dataset ranged from 0.77 to 5.58 km/km^2 (Table 1). The SR, BCP, QPL 90%, and QPL 90% approaches yielded fairly similar threshold estimates (0.69 to 1.10 km/km^2). The QPC 99% and QPL 99% also gave similar threshold estimates (2.04–2.05 km/km^2). Graphically, the INTOL-RDDENS dataset appeared to most closely resemble a wedge-shaped dataset, in which case the QPC 99% and QPL 99% likely did a better job at estimating the disturbance thresholds (Fig. 2). For the INTOL-POPDENS and INTOL-URBAN datasets, estimated disturbance thresholds were similar across all of the

approaches. The estimated thresholds ranged from 12.50 to 16.73 people/km^2 for the INTOL-POPDENS dataset and ranged from 0.82% to 3.10% for the INTOL-URBAN dataset (Table 1). Graphically, there appeared to be a fairly drastic change in percentage intolerant fish at relatively low human population densities and percentage urban land use within watersheds, suggesting that these datasets were more similar to a stair-step disturbance-response threshold pattern. The similarity in the threshold estimates among the approaches for the INTOL-POPDENS and INTOL-URBAN datasets thus is reasonable based on the similarity among the approaches for the simulated stair-step dataset.

For the IBI-RDDENS dataset, estimated disturbance thresholds ranged from 0.34 to 1.81 km/km^2 (Table 1). The BCP and NRD approaches resulted in similar threshold estimates (1.60–1.83 km/km^2), as did the QPC 99% and QPL 99% approaches (Table 1). The SR, QPC 90%, and QPL 90% approaches gave threshold estimates ranging from 0.34 to 0.73 km/km^2, which was substantially lower then the threshold estimates from the other approaches. Graphically, the IBI-RDDENS most closely resembled a

Fig. 3 Scatterplots of the
Michigan fish biological
indicator data in relation to the
human disturbance data.
Locations of the threshold
estimates from each of the
identification methods are also
shown on the plots. INTOL,
percentage of captured fish that
were human disturbance
intolerant (%); IBI, index of
biotic integrity scores;
RDDENS, watershed road
density (km/km^2); POPDENS,
watershed population density
(people/km^2); URBAN,
percentage urban land use
within watershed (%)
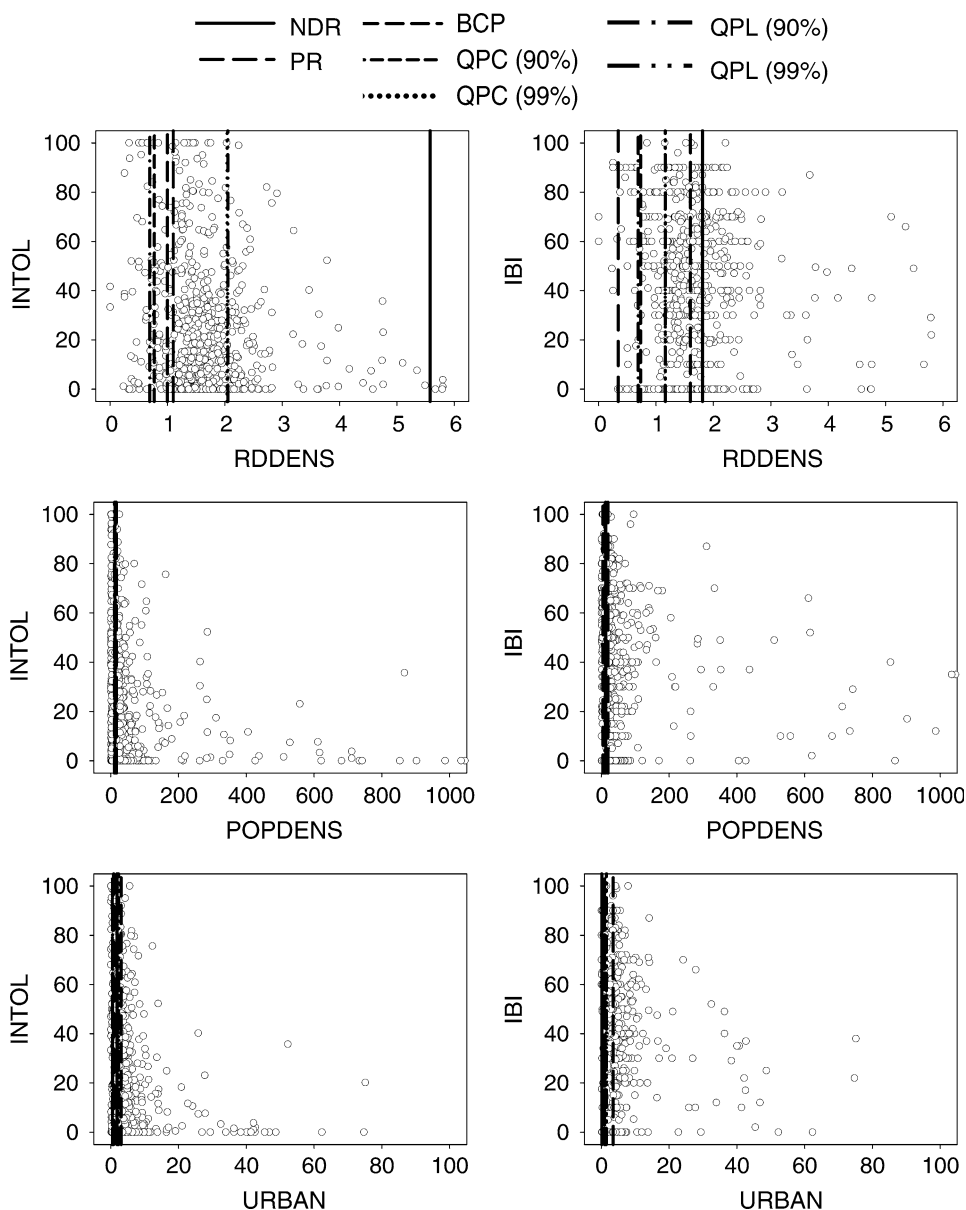
wedge-shaped dataset, and the NDR, BCP, QPC 99%, or
QPL 99% yielded the most likely disturbance threshold
estimates (Fig. 3). For the IBI-POPDENS and IBI-URBAN
datasets, all approaches gave similar threshold estimates
(Table 1). For the IBI-POPDENS dataset, estimated
thresholds ranged from 5.13 to 12.92 people/km^2
(Table 1). For the IBI-URBAN dataset, estimated thresh-
olds ranged from 0.12% to 3.50% (Table 1). Graphically,
all of these threshold estimates appeared reasonable given
the observed disturbance-response variable relationship.

Wisconsin Data

For both the IBI-TP and the IBI-DP datasets, the SR, BCP,
QPC, and QPL approaches gave very similar threshold

estimates, but the estimated threshold from the NDR
approach differed substantially from the other approaches.
For the IBI-TP dataset, the SR, BCP, QPC, and QPL
approaches had threshold estimates that ranged from 0.03
to 0.07 mg/l, while the NDR approach gave a threshold
estimate of 0.39 mg/l. For the IBI-DP dataset, the SR, BCP,
QPC, and QPL approaches gave threshold estimates that
ranged from 0.02 to 0.07 mg/l, while the NDR approach
gave a threshold estimate of 0.30 mg/l. For the IBI-NH4
dataset, the threshold estimates were similar among the
approaches, ranging from 0.02 to 0.07 mg/l. Graphically, it
appeared that there may have been multiple disturbance
thresholds for the IBI-TP, IBI-DP, and IBI-NH4 datasets,
which could explain the differences in the threshold esti-
mates among approaches (Fig. 4).

Fig. 4 Scatterplots of the Wisconsin fish biological indicator data in relation to types and levels of nutrient. Locations of the threshold estimates from each of the identification methods are also shown on the plots. SALMN, number of captured fish that belonged to the Salmonidae family; INTOL, percentage of captured fish that were human disturbance intolerant (%); IBI, index of biotic integrity scores; TP, total phosphorous concentration (mg/l); DP, dissolved phosphorous concentration (mg/l); NH_4, dissolved ammonia nitrogen concentration (mg/l)
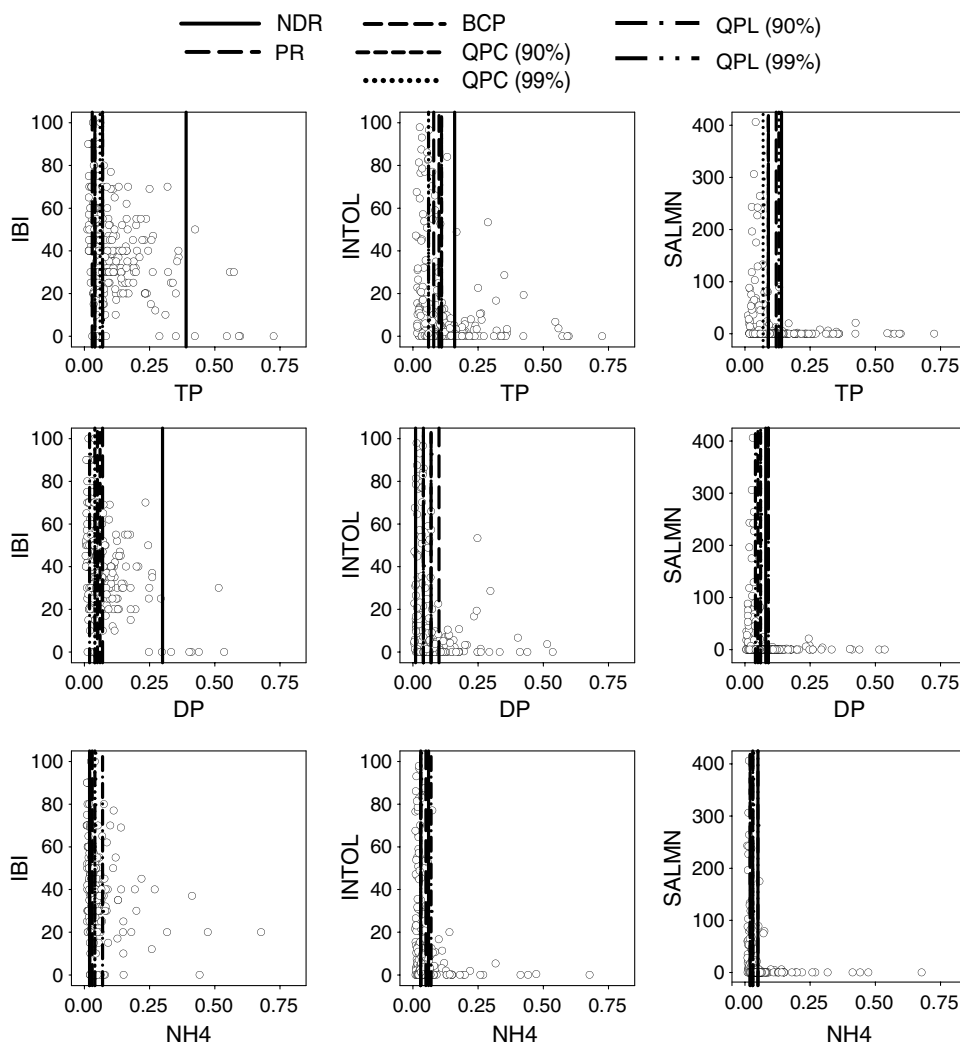
There was very little difference in threshold estimates among the NDR, PR, BCP, QPC, and QPL approaches for the remaining Wisconsin datasets (Table 1). For the INTOL response variable, estimated thresholds ranged from 0.06 to 0.16 mg/L for total phosphorous, 0.01 to 0.10 mg/l for dissolved phosphorous, and 0.03 to 0.07 for dissolved ammonia nitrogen. For the INTOL-TP dataset, it appeared that NDR approach might have overestimated the disturbance threshold (Fig. 4). For the INTOL-DP dataset, the NDR approach may have underestimated the disturbance threshold, while the PR approach may have overestimated the disturbance threshold (Fig. 4). For the SALM-TP dataset, the BCP, QPC 90%, QPC 99%, and QPL 90% approaches appeared to yield the most reasonable threshold estimates, while the NDR, PR, and QPL 99% approaches appeared to slightly overestimate the disturbance threshold. Again, such assessments are complicated by the uncertainty as to whether thresholds actually existed for any of these datasets.

Discussion

There are several advantages in using a quantitative approach to identify disturbance thresholds. First, unlike the simple percentile approach, many quantitative approaches attempt to identify thresholds based on observable changes in response variables, which helps ensure that thresholds correspond to meaningful changes in ecosystem function or biological communities. This property should also help make the approaches relatively robust to natural gradients in intensity of anthropogenic activities. Second, model-based approaches for identifying disturbance thresholds allow estimates of uncertainty in the threshold estimates. For example, with the BCP approach, percentiles of the threshold's marginal posterior distribution can be used to estimate Bayesian credibility intervals for the threshold. With the PR approach, confidence intervals for the estimated threshold can be obtained by bootstrapping or through large-sample approximation. Bootstrapping has also been used to calculate confidence intervals for tree-

based approaches (Qian and others 2003), although caution should be exercised in using bootstrapping with tree-based models as coverage of confidence intervals may be too narrow (Bühlmann and Yu 2002). Finally, quantitative approaches for identifying disturbance thresholds allow testing of the threshold's statistical significance. For example, a chi-square test can be used to test whether thresholds identified by the NDR approach are statistically significant. Both permutation hypothesis tests (Kim and others 2000) and Bayes Information Criteria (BIC) model selection (Tiwari and others 2005) can be used to test the statistical significance of the threshold estimate with the PR approach. Qian and others (2003) describe how to test the significance of a threshold for the BCP approach.

The results of this study suggest that no single method for identifying thresholds necessarily outperforms other methods; accuracy of threshold estimates from the different approaches depends largely on the observed data distributions. Most of the quantitative approaches that we tested performed reasonably well for datasets with drastic changes in response variables, such as the simulated stair-step dataset and INTOL-POPDENS, INTOL-URBAN, IBI-URBAN, SALMN-NH4, INTOL-NH4, and IBI-NH4 data pairs. In contrast, only a few of the approaches, such as the PR and QPL methods, performed well for datasets with gradual decreases in response variables under intensified levels of disturbance, such as the simulated conditional-mean and wedged-shape datasets and the INTOL-RDDENS and IBI-RDDENS datasets. Again, it is important to keep in mind that quantitative or model-based approaches to identifying disturbance thresholds make different assumptions regarding the nature of data distribution, and it is important to pick an approach that assumes a disturbance-response threshold pattern similar to the dataset that is being evaluated. Based on our assessment, one of the more robust approaches for identifying disturbance thresholds was the QPL approach. It performed well for all simulated datasets and it appeared to yield reasonable threshold estimates for many of the field-collected datasets. Thus, if one is uncertain as to the exact nature of a disturbance-response variable relationship, the QPL approach may be a beneficial approach to try initially. However, one can likely imagine a disturbance-response threshold pattern for which the QPL approach would not work well; as a result, it is important not to consider the QPL approach as the ideal threshold identification method that will yield valid estimates under all situations.

Regarding the BCP approach and its performance in identifying thresholds, it is important to note that we limited our consideration of this method to what was described by Qian and others (2003). However, the BCP approach is in fact a far more versatile approach for identifying thresholds because it requires specifying the probability distributions for the different groups. While in this study we limited our use of the approach to relatively simple circumstances, in actuality it would be possible to use the BCP approach to conduct a variety of analyses. For example, the BCP approach could be used to identify disturbance thresholds in a manner similar to the PR approach. Doing so would only require specification of the correct probability distributions for the groups. There are a number of advantages in adopting a Bayesian framework to identify thresholds. As previously mentioned, it allows specification of priors for model parameters, which allows analyses to build on results from previously completed research. A Bayesian framework also results in a more straightforward interpretation of the uncertainty associated with parameters estimates. Perhaps the biggest difficulty in using the BCP approach to identify disturbance thresholds is that many aquatic resource managers may lack the necessary analytic skills to implementing this threshold identification approach. Additional development of the BCP approach for identifying thresholds should be conducted, and we encourage the development of a tool, such as an R package, which would allow managers to use this approach to identify thresholds under a number of possible scenarios (e.g., conditional mean, wedge-shaped disturbance-response threshold patterns).

From our analysis of field-collected datasets, we found that categorizing observed disturbance-response variable patterns into particular distribution types (e.g., stair step, conditional mean, wedge shaped) could be difficult, which is perhaps the biggest challenge in selecting a threshold identification approach. Factors such as small sample sizes, measurement error, and lack of contrast among disturbance measurements can result in dramatically different disturbance-response variable patterns. Because of this, we recommend the following steps be used to identify environmental thresholds. First, the disturbance-response variable relationship should be plotted and visually inspected to help ensure that the dataset includes observation across a broad spectrum of disturbance levels. If the dataset do not span a wide range of disturbance levels, data collection may need to be expanded to additional areas. Visual inspection of plots will also help analysts evaluate whether a single threshold or, possibly, multiple thresholds occur within a dataset. Second, analysts should hypothesize what the relationship between the disturbance and the response variable is. Visual inspection of the plots generated in the previous step will also help with this task, but analysts should not restrict this evaluation to data plots only. Rather, analysts should attempt to incorporate mechanistic understanding of disturbance-response variable relationships. Third, when estimating disturbance thresholds, analysts should consider methods that generate estimates of uncertainty concerning threshold locations and

that allow significance testing regarding the occurrence of environmental thresholds. If using a quantile-based approach, analysts should consider estimating thresholds using multiple percentiles and calculating the rate of change in threshold estimates for these different percentiles. Finally, the location of the threshold estimates should be plotted along with the disturbance-response data to assist in evaluating whether the identified threshold makes sense given the observed data and what already is known about the disturbance and response variable relationship. Based on the results from this last step, it may be necessary to return to an earlier step (e.g., additional data collection, choosing a different threshold identification method) in the process to ensure that the identified thresholds are appropriate.

References

Allan JD (2004) Landscape and riverscapes: the influence of land use on river ecosystems. Annual Review of Ecology and Systematics 35:257–284

Booth DB, Jackson CR (1997) Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. Journal Of The American Water Resources Association 33:1077–1090

Breiman L, Friedman JH, Olshen R, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA

Bühlmann P, Yu B (2002) Analyzing bagging. Annals Of Statistics 30:927–961

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1:412–420

Cade BS, Terrell JW, Schroeder RL (1999) Estimating effects of limiting factors with regression quantiles. Ecology 80:311–323

Chaudhuri P, Loh WY (2002) Nonparametric estimation of conditional quantiles using quantile regression trees. Bernoulli 8:561–576

Denoël M, Ficetola GF (2007) Landscape-level thresholds and newt conservation. Ecological Applications 17:302–309

EPA (U.S. Environmental Protection Agency) (2000a) Ambient water quality criteria recommendations, information supporting the development of state and tribal nutrient criteria, rivers and streams in nutrient ecoregions VII and VIII. EPA 822-B-00-017. EPA, Office of Water, Washington, DC

EPA (U.S. Environmental Protection Agency) (2000b) Mid-Atlantic highland streams assessment. EPA903/R-00/015 822-B-00-017. EPA, Region 3, Washington, DC

Fore LS, Karr JR, Wisseman RW (1996) Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal Of The North American Benthological Society 15:212–231

Groffman PM, Baron JS, Blett T, Gold AJ, Goodman I, Gunderson LH., Levinson BM, Palmer MA, Paerl HW, Peterson GD, Poff NL, Rejeski DW, Reynolds JF, Turner MG, Weathers KC, Wiens J (2006) Ecological thresholds: the key to successful environmental management or an important concept with no practical application? Ecosystems 9:1–13

Karr JR, Chu EW (1999) Restoring life in running waters, better biological monitoring. Island Press, Covelo, CA

Kim HY, Fay MP, Feuer EJ, Midthune DN (2000) Permutation methods for joinpoint regression with applications to cancer rates. Statistics In Medicine 19:335–351

King RS, Richardson CJ (2003) Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. Environmental Management 31:795–809

Lerman PM (1980) Fitting segmented regression models by grid search. Applied Statistics 29:77–84

Loh WY (2002) Regression trees with unbiased variable selection and interaction detection. Statistica Sinica 12:361–386

Loh WY (2007) GUIDE (version 5.2) User manual. University of Wisconsin, Madison. Available at: http://www.stat.wisc.edu/~loh/treeprogs/guide/guideman.pdf. Accessed October 15, 2007

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337

Lyons J (1992) Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. General Technical Report NC-149. U.S. Department of Agriculture, Forest Service, North Central Experiment Station, St. Paul, MN

Lyons J, Wang L, Simonson TD (1996) Development and validation of an index of biotic integrity for cold-water streams in Wisconsin. North American Journal of Fisheries Management 16:241–256

Muggeo VMR (2003) Estimating regression models with unknown break-points. Statistics in Medicine 22:3055–3071

National Cancer Institute (2005) Joinpoint regression program, version 3.0. Silver Spring, Maryland. Available at: http://srab.cancer.gov/joinpoint. Accessed October 15, 2007

Paul MJ, Meyer L (2001) Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333–365

Qian SS, King RS, Richardson CJ (2003) Two statistical methods for the detection of environmental thresholds. Ecological Modelling 166:87–97

Qian SS, Pan Y, King RS (2004) Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. Ecological Indicators 4:29–37

R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

SAS Institute (2003) SAS/STAT user's guide. SAS Institute, Cary, NC

Schueler T (1994) The importance of imperviousness. Watershed Protection Techniques 1:100–111

Seber GAF, Wild CJ (2003) Nonlinear regression. John Wiley & Sons, Hoboken, NJ

Tiwari RC, Cronin KA, Davis W, Feuer EJ, Yu B, Chib S (2005) Bayesian model selection for join point regression with application to age-adjusted cancer rates. Applied Statistics 54:919–939

Therneau TM, Atkinson B, Ripley B (2007) rpart: recursive partitioning. R package version 3.1-36. Available at: http://cran.r-project.org. Accessed October 15, 2007

Toms JD, Lesperance ML (2003) Piecewise regression: a tool for identifying ecological thresholds. Ecology 84:2034–2041

Wang L, Lyons J (2003) Fish and benthic macroinvertebrate assemblages as indicators of stream degradation in urbanizing watersheds. In: Simon TP (ed) Biological response signatures: multimetric index patterns for assessment of freshwater aquatic assemblages. CRC Press, Boca Raton, FL

Wang L, Lyons J, Kanehl P (2003) Impacts of urban land cover on trout streams in Wisconsin and Minnesota. Transactions of the American Fisheries Society 132:825–839

Wang L, Robertson DM, Garrison PJ (2007) Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. Environmental Management 39:194–212

Wang L, Seelbach PW, Hughes RM (2006a) Introduction to landscape influences on stream habitats and biological assemblages. American Fisheries Society Symposium 48:1–23

Wang L, Seelbach PW, Lyons J (2006b) Effects of levels of human disturbance on the influence of watershed, riparian, and reach scale factors on fish assemblages. American Fisheries Society Symposium 48:199–219

Wang L, Brenden T, Seelbach PW, Cooper A, Allan D, Clark R Jr, Wiley M (2008) Landscape based identification of human disturbance gradients and references for streams in Michigan. Environ Monitor Assess 141:1–17