

Empirical Approaches for Nutrient Criteria Derivation

**Prepared by:
United States Environmental Protection Agency,
Office of Water, Office of Science and Technology**

**Science Advisory Board Review Draft
August 17, 2009**

CONTRIBUTORS

USEPA, Office of Research and Development, National Center for Environmental Assessment, Washington, DC

Lester L. Yuan*

USEPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Research Triangle Park, NC

John F. Paul*

United States Environmental Protection Agency (USEPA), Office of Water, Office of Science and Technology, Washington DC

Robert Cantilli

Ifeyinwa Davis

Jim Keating

Dana Thomas

Technical Support – Tetra Tech, Inc., Center for Ecological Sciences, Owings Mills, MD

Michael J. Paul*

*denotes primary author

DISCLAIMER

This document provides technical guidance to States, authorized Tribes, and other authorized jurisdictions to develop water quality criteria and water quality standards under the Clean Water Act (CWA) to protect against the adverse effects of nutrient over-enrichment. Under the CWA, States and authorized Tribes are to establish water quality criteria to protect designated uses. State and Tribal decision-makers retain the discretion to adopt approaches on a case-by-case basis that differ from this guidance when appropriate. While this document presents methods to strengthen the scientific foundation for deriving nutrient criteria, it does not substitute for the CWA or USEPA regulations; nor is it a regulation itself. Thus it cannot impose legally binding requirements on USEPA, States, authorized Tribes, or the regulated community, and it might not apply to a particular situation or circumstance. USEPA may change this guidance in the future.

ACKNOWLEDGMENTS

The authors wish to gratefully acknowledge key internal reviewers whose efforts and input greatly improved earlier drafts of this document: Mark Barath (USEPA Region 3), Susan Cormier (USEPA ORD), James Curtin (USEPA OGC), Christopher Day (USEPA Region 3), Terry Fleming (USEPA Region 9), John Goodin (USEPA OWOW), Treda Grayson (USEPA OWOW), Lareina Guenzel (USEPA Region 8), Michael Haire (USEPA OWOW), John Hochheimer (Tetra Tech Inc.), Neil Kamman (Vermont Department of Environmental Conservation), John Kelly (USEPA ORD), Ryan King (Baylor University), Tina Laidlaw (USEPA Region 8), Peter Leinenbach (USEPA Region 10), Dave Mount (USEPA ORD), Susan Norton (USEPA ORD), Barbara Pace (USEPA OGC), Maggie Passmore (USEPA Region 3), Suesan Saucerman (USEPA Region 9), John van Sickle (USEPA ORD), Qian Song (Duke University), Danny Wiegand (USEPA ORD/OSP) and Izabela Wojtenko (USEPA Region 2).

CONTENTS

| | |
|---|------------|
| ACKNOWLEDGMENTS | iv |
| LIST OF FIGURES..... | vii |
| LIST OF TABLES..... | x |
| EXECUTIVE SUMMARY | xi |
| Introduction..... | 1 |
| <i>Overview of nutrient criteria derivation approaches.....</i> | <i>1</i> |
| Reference condition approach..... | 2 |
| Stressor-response approach | 2 |
| Mechanistic modeling approach..... | 2 |
| <i>Purpose of this document</i> | <i>2</i> |
| <i>A five step process for using stressor-response relationships to derive nutrient criteria.....</i> | <i>3</i> |
| 1 Selecting and exploring stressor and response variables | 5 |
| 1.1 <i>Selecting stressor and response variables.....</i> | <i>5</i> |
| 1.2 <i>Exploring, visualizing, and summarizing available data</i> | <i>6</i> |
| 1.3 <i>Basic data visualization.....</i> | <i>6</i> |
| 1.4 <i>Maps</i> | <i>11</i> |
| 1.5 <i>Conditional probability.....</i> | <i>12</i> |
| 1.6 <i>Classifications.....</i> | <i>13</i> |
| 2 Assessing the strength of the cause-effect relationship | 14 |
| 2.1 <i>Conceptual models.....</i> | <i>15</i> |
| 2.2 <i>Existing literature</i> | <i>15</i> |
| 2.3 <i>Alternate methods for assessing causal effects.....</i> | <i>16</i> |
| 3 Analyzing data | 17 |
| 3.1 <i>Translating a response threshold to a candidate criterion: single explanatory variable</i> | <i>18</i> |
| 3.1.1 <i>Simple linear regression (SLR)</i> | <i>20</i> |
| 3.1.2 <i>Quantile regression (QR)</i> | <i>25</i> |
| 3.1.3 <i>Logistic regression</i> | <i>28</i> |
| 3.1.4 <i>Application of bivariate models</i> | <i>29</i> |
| 3.2 <i>Translating a response threshold to a candidate criterion: multiple explanatory variables.....</i> | <i>30</i> |
| 3.2.1 <i>Multiple linear regression (MLR).....</i> | <i>30</i> |
| 3.3 <i>Identifying thresholds from stressor-response relationships</i> | <i>35</i> |
| 3.3.1 <i>Non-parametric changepoint analysis (nCPA).....</i> | <i>36</i> |
| 3.3.2 <i>Discontinuous regression models</i> | <i>39</i> |
| 4 Evaluating estimated stressor-response relationships | 40 |

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

| | | |
|----------|--|-----------|
| 4.1 | <i>Model validation</i> | 41 |
| 4.2 | <i>Qualitative assessment of the uncertainty of the estimated stressor-response relationship</i> | 43 |
| 4.3 | <i>Select the stressor-response model</i> | 44 |
| 5 | Evaluating candidate stressor-response criteria | 46 |
| | References | 53 |
| | Appendix A – Data Descriptions | 62 |
| | Appendix B – Propensity Score Analysis | 64 |
| | Appendix C – Conditional Probability Analysis | 67 |
| | Appendix D – Acronyms | 74 |
| | Appendix E – Glossary | 75 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1 – Detailed flow diagram of steps in stressor-response relationship based nutrient criteria derivation. | 4 |
| Figure 2. Examples of histograms of EMAP-West Streams Survey log-transformed TN and TP. | 6 |
| Figure 3. Example boxplots from EMAP-West Streams Survey data for log(TN)(left plot) and total species richness (right plot)..... | 7 |
| Figure 4. Quantile-quantile plots comparing TN (left plot) and log(TN) (right plot) values from the EMAP-West Streams Survey to normal distributions..... | 8 |
| Figure 5. Cumulative distribution functions for phosphorus from EMAP Northeast Lakes Survey data | 9 |
| Figure 6. Scatter plot of log(TP) vs. total richness in the EMAP-West Streams Survey dataset. | 9 |
| Figure 7. Scatter plot matrix of EMAP-West Streams Survey TN and TP (as log-transformed variables) against measures of animal grazing in the watershed, percent sand/fine substrates and total macroinvertebrate richness | 11 |
| Figure 8. Map of EMAP Northeast Lakes Survey sample sites | 12 |
| Figure 9. Examples of exploratory data analysis plots using conditional probability analysis (CPA) of EMAP Northeast Lakes Survey data for chlorophyll <i>a</i> as a response variable (threshold at 15 µg/L) and potential stressor variables of log(TP), log(TN), dissolved organic carbon (DOC), and turbidity..... | 13 |
| Figure 10. Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts | 16 |
| Figure 11. Plot showing example of stressor-response relationship, where the response is a direct measure of designated use or can be easily linked to a designated use measure. | 19 |
| Figure 12. Log(TP) versus log(chl <i>a</i>) using EMAP Northeast Lakes Survey data | 22 |
| Figure 13. Log(TN) versus total species richness in EMAP-West Xeric region streams... .. | 24 |
| Figure 14. Log(TP) versus EPT species richness using the least urban MBSS Piedmont ecoregion stream dataset | 25 |
| Figure 15. Log(TP) vs. log(chl <i>a</i>) for EMAP Northeast Lakes Survey | 26 |
| Figure 16. Log(TN) vs. total species richness for EMAP-West Xeric region streams | 27 |
| Figure 17. Log(TP) vs. benthic diatom TSI for the EMAP-MAHA/USGS-NAWQA regional dataset | 28 |

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Figure 18. Modeled probability of impairment (i.e., total richness < 40) as a function of log TN for EMAP-West Xeric region streams. 29

Figure 19. Log(TN) versus percent sand/fines in EMAP-West Xeric region streams..... 32

Figure 20. Log(TP) versus log(TN) for EMAP Northeast Lakes Survey sites..... 34

Figure 21. Log(TP) versus epifaunal substrate scores from the least urban MBSS Piedmont ecoregion stream dataset 35

Figure 22. Stressor-response relationship showing threshold behavior 35

Figure 23. Hypothetical relationship between an independent variable X and a dependent variable Y 36

Figure 24. Plot of chlorophyll *a* versus log total phosphorus in EMAP Northeast Lakes Survey sites 38

Figure 25. Log(TN) vs. total species richness in EMAP-West Xeric region streams..... 39

Figure 26. Log(TP) vs. total species richness for EMAP-West Stream Survey data 40

Figure 27. Predicted log(chl *a*) vs. observed log(chl *a*) for EMAP Northeast Lakes Survey sites in Massachusetts 42

Figure 28. Simple linear regression estimate of stressor-response relationship 46

Figure 29. Example of using an estimated stressor-response relationship to project the effects of criterion. 47

Figure 30. Example of projecting chlorophyll *a* in lakes to a candidate criterion value of log(TP) = 2 (TP = 100 µg/L). 47

Figure 31. Resulting distribution of chlorophyll *a* after projection 48

Figure 32. Observed and predicted probabilities of impairment for different nutrient concentrations 49

Figure 33. Total species richness vs. log(TN) for EMAP-West Xeric region streams. 50

Figure 34. Observed and predicted probabilities of impairment for EMAP-West Xeric region streams 50

Figure 35. Relationships between TN and total invertebrate richness within each stratum..... 66

Figure 36. CPROB add-in module for conducting conditional probability analysis. 70

Figure 37. Cumulative distribution function (CDF) and conditional cumulative distribution functions (CCDFs) for total phosphorus concentrations for the EMAP Northeast Lakes Survey data 71

Figure 38. The empirical conditional probability curve (black open circles) for observing a eutrophic lake (chlorophyll *a* > 15 µg/L) if the total phosphorus concentration in the lake exceeds the x-axis value 72

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Figure 39. Cumulative distribution function (CDF) and conditional cumulative distribution functions (CCDFs) for EMAP-West Xeric region streams 73

Figure 40. The empirical conditional probability curve (black open circles) for observing a stream with total richness < 40 if the total nitrogen concentration in the stream exceeds the x-axis value. 73

LIST OF TABLES

Table 1. Summary of different method characteristics..... 44
Table 2. Correlation coefficients between covariates and TN..... 65

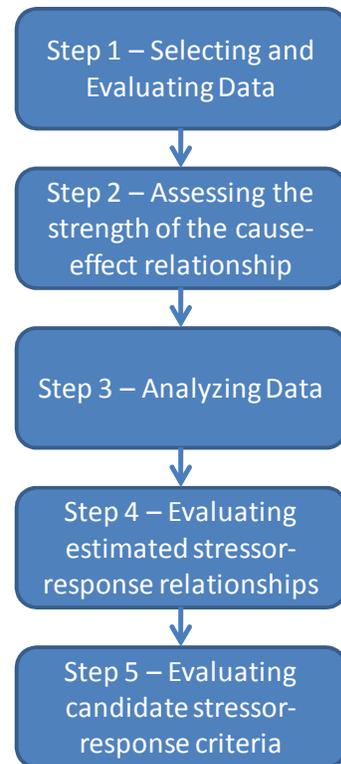
EXECUTIVE SUMMARY

As prescribed by the Clean Water Act, numeric nutrient water quality criteria are derived to protect the designated uses of waterbodies from nutrient over-enrichment. The U.S. Environmental Protection Agency (USEPA) recommends three types of approaches for setting numeric nutrient criteria (USEPA 2000a and 2000b), including: reference condition approaches, stressor-response analysis, and mechanistic modeling. This document elaborates on the second of these three and its purpose is to provide information on the scientific foundation for using empirical approaches to describe stressor-response relationships for deriving numeric nutrient criteria. This document describes a five step process for analyzing stressor-response relationships that can be used to derive numeric nutrient criteria.

The document begins with a brief introduction to nutrient criteria and the criteria derivation approach and reviews the purpose and scope. This section introduces the process by which stressor-response relationships can be used to derive nutrient criteria. This process is laid out in five steps.

The first section of the document describes step 1, selecting and evaluating data. Step 1 reviews techniques for selecting the variables that appropriately quantify the stressor (i.e., excess nutrients) and the response (e.g., chlorophyll *a* (chl *a*), dissolved oxygen, or a biological index). Selecting response variables that relate directly to measures of designated use are most appropriate since criteria must ensure protection of the designated uses. This step then describes data exploration, visualization, and summary. Exploratory techniques include histograms, box and whisker plots, Quantile-Quantile plots, cumulative distribution plots, scatter diagrams, and spatial mapping. The visualization step helps the analyst understand how variables change across space and time, general relationships among variables and how one or more variables co-vary. Conditional probability analysis, a more quantitative statistical approach for summarizing the data can also be used for data exploration.

Step 2 is assessing the strength of the cause-effect relationship represented in the stressor-response linkage. When stressor-response relationships are used to establish nutrient criteria, it is important to assess the degree to which changes in nutrient concentration are likely to cause changes in the chosen response variable. This can be accomplished using conceptual models, existing literature, and empirical models. The



Empirical Approaches for Nutrient Criteria Derivation SAB Review Draft

step 2 section describes these approaches and their application for nutrient criteria derivation.

Step 3 is data analysis. In Step 3 of the process, data are analyzed to estimate stressor-response relationships. Types of analyses are presented in two main subsections that provide different approaches for deriving criteria: (1) translating existing response thresholds into candidate criteria using estimated stressor-response relationships and (2) identifying thresholds from inherent characteristics of the stressor-response relationships. Each subsection begins with an overview of the process by which an estimated stressor-response relationship is used to derive a nutrient criterion with each approach. Then, different statistical methods are described, the data requirements for each are reviewed, and example applications relevant to nutrient criteria derivation are presented. The first part of the step 3 section focuses on both bivariate modeling techniques including linear regression, logistic regression, and quantile regression, as well as multiple linear regression for use with multiple predictors. Regression techniques come with underlying assumptions that, if met, provide a relatively easy way to analyze data and evaluate stressor-response relationships for deriving candidate criteria. When there are multiple factors contributing to a response, simple linear models often can be strengthened by incorporating additional predictors or stratifying on other variables. Multiple linear regression models are one approach for doing this. Multiple regression models require more data, but they provide an equation that shows the relationship between two or more stressors and a response simultaneously. Different considerations for using particular bivariate and multiple regression models are discussed. The second part of the step 3 section describes techniques for identifying a threshold in a stressor-response relationship that may be used to derive criteria. At times, the relationship of response variables to nutrients changes as nutrient concentrations increase. Nutrient concentrations associated with this change in response provide another way to derive a criterion. Non-parametric changepoint analysis and discontinuous regression modeling methods are presented as techniques to identify these changes or thresholds.

Step 4 in the process is evaluating the estimated stressor-response relationships. A number of approaches can be used to develop estimates of stressor-response relationships, and so this step reviews how to evaluate and compare different estimates of the stressor-response relationship and select those most appropriate for criteria derivation. The step 4 section reviews the validation of predictive performance for a model. The final objective in step 4 is to demonstrate how to select a stressor-response model using the response variable that best represents the data.

The final step in the process, step 5, is evaluating candidate stressor-response criteria. The process of using a stressor-response relationship to translate an existing response threshold to a comparable nutrient criterion is designed so that acceptable values of response variables are *predicted* at lower nutrient concentrations. That is, a criterion should be set to the upper-most value of nutrient concentrations that are likely to result

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

in response variables that meet management objectives (e.g., protect designated uses). In this last step of the process, an approach is outlined for predicting conditions after implementing different nutrient criteria. These predictions can then be used to evaluate different candidate criteria and guide the selection of an optimal value in terms of resource protection based on the stressor-response relationships.

Nutrient criteria are important for protecting our nation's waterbodies from the negative effects of excessive nutrients. These criteria can be derived using a variety of approaches, including stressor-response relationships. This guidance lays out a specific process for conducting such analyses. It builds off existing nutrient criteria guidance and provides further scientific support for using empirical approaches to derive nutrient criteria using stressor-response relationships. The process described will support States, Territories, and authorized Tribes in incorporating stressor-response relationships into their nutrient criteria development programs and further the goal of reducing nutrient pollution nationwide.

Introduction

Under the Clean Water Act (CWA), States, Territories, and authorized Tribes are responsible for establishing water quality standards (WQS) that include designated uses such as aquatic life and recreation, criteria that define levels of water quality variables protective of the designated uses, and an antidegradation provision. Nutrients are essential for plant and microbial growth and at natural concentrations are generally considered beneficial. Over-enrichment by nitrogen and phosphorus stimulates excessive rates of plant and microbial growth and can produce biological and physical responses in surface water that adversely affect water quality and aquatic life. Nutrients affect both waterbodies directly receiving nutrients and downstream waterbodies. The United States Environmental Protection Agency (USEPA) nutrient criteria guidance recommends that criteria be derived for both total nitrogen (TN) and total phosphorus (TP) (primary causal variables) and chlorophyll *a* and clarity (primary response variables) that are protective of designated uses. The guidance does not preclude the use of alternative causal or response variables, and suggest several additional variables such as dissolved oxygen, trophic state indices, and biocriteria (USEPA 2000a).

Overview of nutrient criteria derivation approaches

The USEPA has provided a series of peer-reviewed technical approaches and methods for deriving nutrient criteria (USEPA 2000a, 2000b, 2001, and 2008). USEPA's view is that the criteria derivation process for the toxic effect of chemical pollutants is not applicable for nutrients because effects, while linked to widespread and significant aquatic degradation, occur through a process of intermediate steps that cannot be easily tested in simple laboratory studies. As a result, nutrient criteria derivation relies in large part on empirical analysis of field data.

Basic analytical approaches for nutrient criteria derivation include: (1) the reference condition approach, (2) stressor-response analysis, and (3) mechanistic modeling (USEPA 2000a, 2000b). The application and consideration of established (e.g., published) nutrient response thresholds is another acceptable approach (USEPA 2000a). Each approach alone may be appropriate for a State to derive scientifically defensible numeric nutrient criteria. As each approach has different characteristics and data needs, these should be considered in the context of individual State situations and available information.

States and USEPA should carefully consider the strengths and characteristics of each analytical approach with respect to data availability and designated use protection needs. As mentioned, threshold values from or based upon published literature provide an additional source of information. Evaluating criteria estimated from

Empirical Approaches for Nutrient Criteria Derivation SAB Review Draft

different approaches can reduce some of the uncertainties inherent in any single approach. Thus, USEPA guidance has made it clear that these approaches may be used either independently or in combination depending upon available data and the context being addressed (USEPA 2000a). The approaches are briefly summarized below.

Reference condition approach

The reference condition approach described in detail in all of the previously published nutrient criteria guidance documents (e.g., USEPA 2000a, 2000b, 2001, and 2008) derives candidate criteria from distributions of nutrient concentrations and biological responses in reference waterbodies. Reference waterbodies (within a waterbody class; e.g., rivers or lakes) represent least disturbed and/or minimally disturbed conditions (sensu Stoddard et al. 2006) and share similar characteristics to the waterbodies for which criteria are being derived. Criteria are derived by compiling measurements of causal and response variables from reference waterbodies and selecting a value from the distribution. USEPA's nutrient criteria guidance recommends the use of percentiles derived from the reference waterbody distributions, since these waterbodies represent an example of the biological integrity expected for a region (USEPA 2000a).

Stressor-response approach

The stressor-response approach involves estimating a relationship between nutrient concentrations and biological response measures related to designated use of a waterbody (e.g., a biological index or recreational use measure) either directly or indirectly, but ideally quantitatively. Then, nutrient concentrations protective of designated uses can be derived from the estimated relationship. This approach relates nutrient concentrations to response measures and thus to designated uses and is the focus of this document.

Mechanistic modeling approach

The mechanistic modeling approach predicts the effects of changes in nutrient concentrations using site-specific parameters and equations that represent ecological processes. Mechanistic models include a wide range of process-based water quality models that are described in some detail in the nutrient criteria guidance documents (e.g. USEPA 2000a, 2000b), and in greater detail in water quality modeling textbooks (e.g., Chapra 1997). Mechanistic models can account for site-specific effects of nitrogen and phosphorus pollution and mechanistically link changes in nutrient concentration to impacts on water quality relevant to biological conditions.

Purpose of this document

As noted, the use of nutrient stressor-response relationships to derive nutrient criteria is one of the recommended approaches in USEPA nutrient criteria guidance (USEPA 2000a,

2000b). The nutrient criteria guidance states that “relations (predictive models) between biocriteria and nutrients could be used to set nutrient and biocriteria, based on a desired level of biotic integrity or other valued ecosystem component” (USEPA 2000a, and see USEPA 2000b, 2008). In addition, USEPA has used stressor-response relationships for other criteria derivation. The Framework for Developing Suspended and Bedded Sediments (SABS) Water Quality Criteria (USEPA 2006a) provides guidance on the use of stressor-response relationships using field-derived data to derive suspended and bedded sediment criteria.

The purpose of this empirical approaches guidance document is to provide information on the scientific foundation for using empirical approaches to describe stressor-response relationships for deriving nutrient criteria. This document supports and is consistent with the existing nutrient criteria guidance (USEPA 2000a, 2000b, 2001, and 2008). The statistical and analytical approaches described below represent a scientifically defensible means of identifying patterns and relationships in field data. Examples illustrate how these approaches could be applied for purposes of deriving criteria. Although the examples provided focus on streams and lakes, the information presented in this document should be applicable to any waterbody type. The document is intended for State, local and tribal government water resource managers and other interested stakeholders, with some scientific training. Information in this document can be used when preparing to evaluate empirical data as well as during the planning phase of a field study or monitoring program. This document is not intended to provide exhaustive methods on how to complete individual analyses, and interested readers should consult qualified analytical experts or appropriate literature for more information on methodological detail.

A five step process for using stressor-response relationships to derive nutrient criteria

Five steps are involved when stressor-response relationships are used to derive numerical nutrient criteria (Figure 1). First, data are assembled, and the nutrient and response variables on which the analysis will focus are selected. Second, the strength of the cause-effect relationship between the selected nutrient and response variables is assessed. Third, data are analyzed to estimate stressor-response relationships, and these stressor-response relationships are used to derive candidate nutrient criteria. Fourth, stressor-response relationships estimated by different statistical approaches are compared and evaluated. Finally, candidate nutrient criteria are evaluated, and appropriate criterion values identified. The next five sections of this document correspond to each of these steps.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

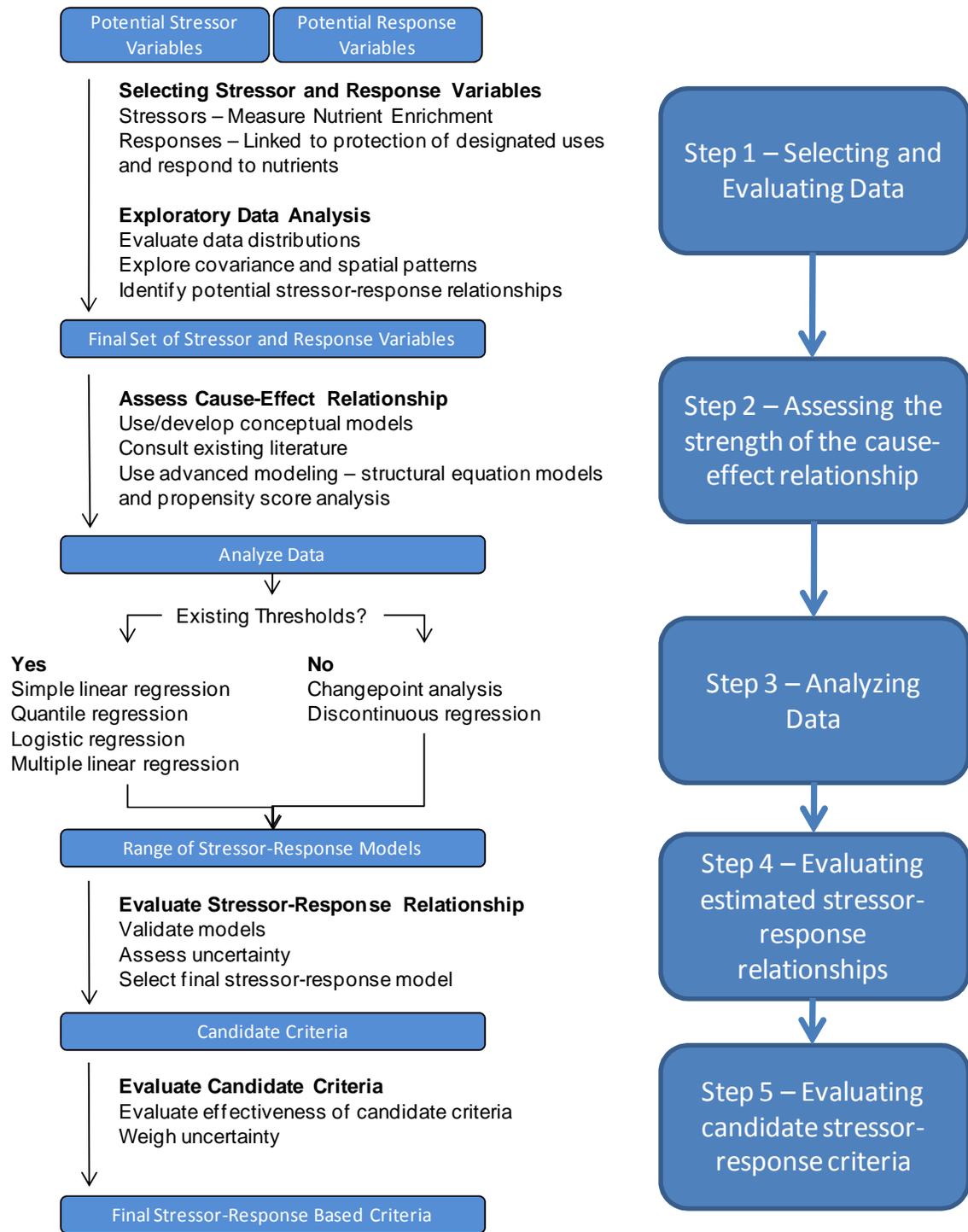
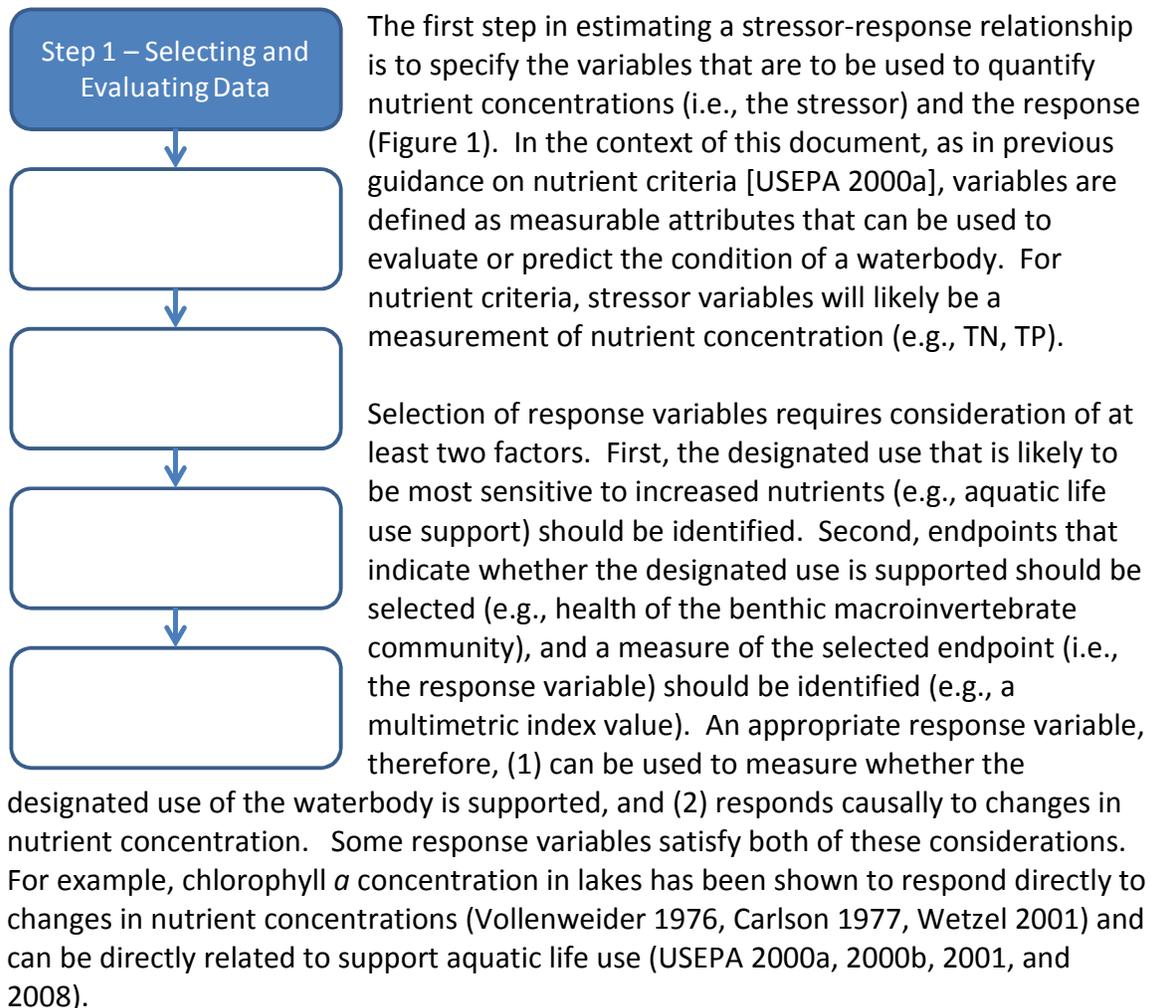


Figure 1 – Detailed flow diagram of steps in stressor-response relationship based nutrient criteria derivation.

1 Selecting and exploring stressor and response variables

1.1 Selecting stressor and response variables



The degree to which a selected response variable satisfies the factors listed above can influence the statistical analysis applied to the data. A response variable that is strongly and causally related to nutrients often may be effectively modeled with a bivariate approach, whereas a response variable that is causally related to nutrients but whose effect may be obscured by other factors may require a model with multiple explanatory variables to accurately estimate the effects of nutrients. In cases in which the selected response variable does not directly measure attainment of the designated use, additional modeling may be required to associate the selected response with use attainment.

1.2 Exploring, visualizing, and summarizing available data

A first step for any analysis of observational field data is to visually examine the data by viewing distributions of and relationships among the variables. This exploratory data analysis (EDA) (Tukey 1977) helps the analyst understand the properties of different variables, which can be useful for determining the validity of applying some of the statistical methods. EDA also provides insights into how different factors co-vary in the dataset, providing a basis for selecting variables to include in subsequent statistical models. In this section, several basic plotting techniques are presented, followed by a description of conditional probability analysis, a more formalized statistical approach for summarizing how changes in nutrient concentrations are associated with the probability of waterbodies attaining their designated use.

1.3 Basic data visualization

The basic visualization techniques that will be discussed are histograms, box and whisker plots, Quantile-Quantile (Q-Q) plots, cumulative distribution functions (CDFs), and scatter plots.

A histogram summarizes the distribution of the data by placing observations into intervals (also called classes or bins) and counting the number of observations in each interval. The y-axis can be number of observations, percent of total, fraction of total, or density (height of bar times width is fraction of observations in the interval). The distribution of bars in the histogram depends upon how one decides to divide the data into intervals. Examples of histograms are shown in Figure 2 for TP and TN from the EMAP-West Streams Survey dataset (Appendix A).

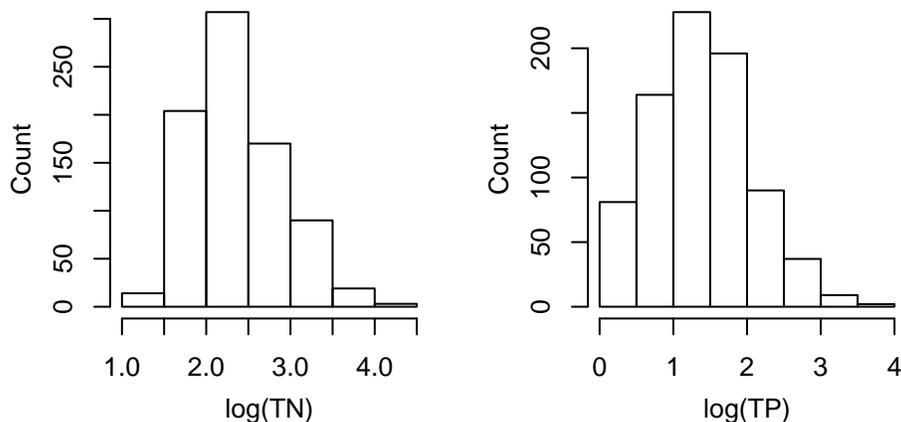


Figure 2. Examples of histograms of EMAP-West Streams Survey log-transformed TN and TP. Units are $\mu\text{g/L}$.

A box and whisker plot (also referred to as boxplot) summarizes the distribution of a variable in terms of 5 numbers: the median (50th percentile), hinges (25th and 75th percentiles), and extremes (minimum and maximum) (Tukey 1977). A standard boxplot

consists of a box defined by the hinges, a line or point on the box at the median, and a line (whisker) drawn from each hinge to the extreme value.

A slight variation on the standard boxplot is shown in Figure 3 for variables from the EMAP-West Stream Survey, where the whiskers extend to a set distance from the hinge, and outliers beyond the span are identified. The span is typically defined as $1.5 \times$ (upper hinge value – lower hinge value or inter-quartile range).

Because of their compact form, boxplots are particularly useful for comparing variable distributions of different data subsets. In Figure 3, the difference in the distributions of TN and total richness across ecoregions can be easily discerned. For example, the Plains region generally has higher total nitrogen concentrations and the mountains generally more macroinvertebrate taxa. One limitation for boxplots is that all of the variables are equally weighted. Hence, sample weights provided by a probability survey cannot be incorporated in this representation of the variable distribution.

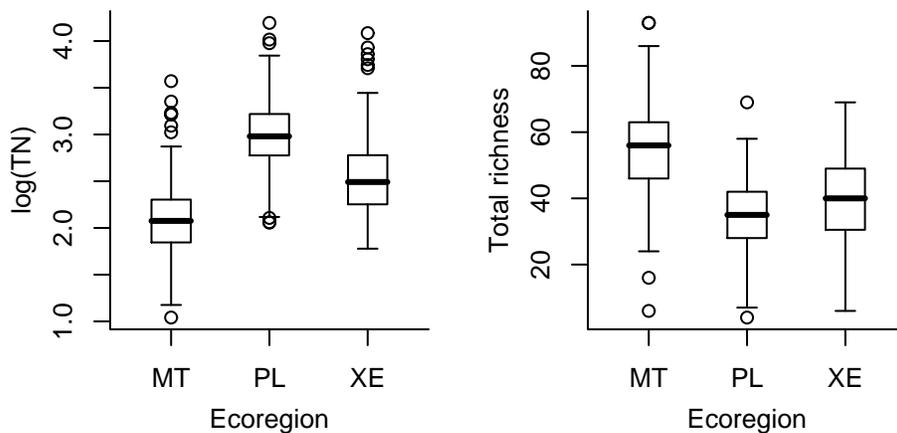


Figure 3. Example boxplots from EMAP-West Streams Survey data for log(TN)(left plot) and total species richness (right plot). Variable distributions within different ecoregions shown. Units are $\mu\text{g/L}$ for Log(TN). MT : Mountains, PL: Plains, XE: Xeric.

A Q-Q plot or probability plot typically compares the distribution of a variable to a particular theoretical distribution (e.g., normal distribution, Wilk and Gnanadesikan 1968). The Q-Q comparison plot with a normal distribution provides a graphical means of assessing the normality of a variable. Q-Q plots for TN values in EMAP-West Streams Survey data indicate that log-transformed values are more normally distributed than raw measurements (Figure 4).

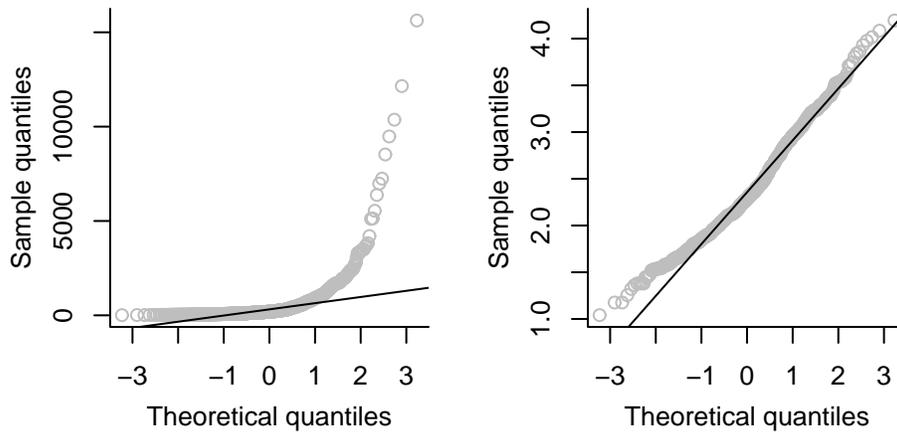


Figure 4. Quantile-quantile plots comparing TN (left plot) and log(TN) (right plot) values from the EMAP-West Streams Survey to normal distributions. Units are $\mu\text{g/L}$.

An empirical CDF is another visual representation of the distribution of values for a particular variable. Possible values of the variable, x , are each plotted against the probability that observations of the variable are less than the specified value. The reverse CDF is commonly used, which displays the probability that the observations are greater than a specified value. In constructing a CDF, weights (inclusion probabilities) can be used if they are available from a probability survey design. In this way the probability that a value of the variable in the statistical population is less than a specified value is estimated. Confidence limits can be placed around the CDF, computed for probability samples by an estimation formula or by resampling. CDFs for phosphorus data from the EMAP Northeast Lakes Survey (Appendix A) are shown in Figure 5. The CDF for the sampled sites is shown in black (unweighted data), while the estimated CDF for the statistical population of all lakes in the northeast U.S. (weighted by inclusion probabilities) is shown in blue. The median phosphorus concentration is that concentration on the x-axis where a line drawn from the proportion value of 0.5 on the y-axis intercepts each cumulative distribution line. For all of the lakes sampled (black line) the value is $11 \mu\text{g/L}$, while the estimated median for all of the lakes in the northeast U.S. (blue line) is $17 \mu\text{g/L}$.

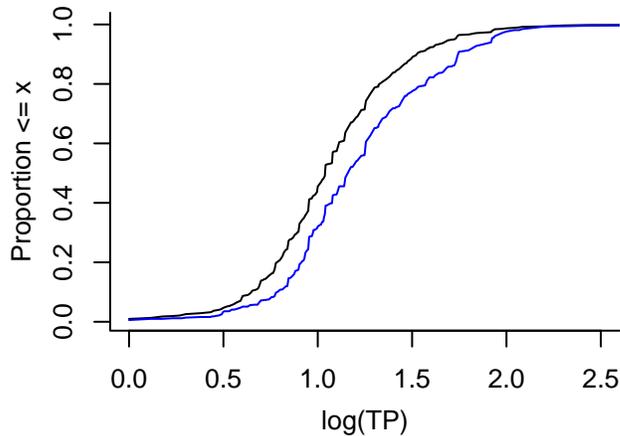


Figure 5. Cumulative distribution functions for phosphorus from EMAP Northeast Lakes Survey data. Black line: unweighted, blue line: weighting as prescribed by the probability design. Units are $\mu\text{g/L}$.

Scatter plots are used to visualize relationships between different variables. An example scatter plot of \log_{10} transformed TP versus total macroinvertebrate species richness is shown in Figure 6. In addition to scatter plots, a useful method for identifying and visualizing a possible underlying relationship between the two plotted variables (independent of any specific model) is to approximate this relationship using a data smoother. Different smoothing algorithms have been proposed, including locally weighted regression (e.g., loess, Cleveland 1979, 1993, Cleveland and Devlin 1988, Cleveland et al. 1992), and penalized regression splines (Wood and Augustin 2002). Most statistical software packages provide access to one or more of these smoothers. The smoothed relationship between TP and total species richness in this example suggests that a threshold exists at approximately $\text{TP} = 10 \mu\text{g/L}$ ($\log[\text{TP}] = 1$). Total richness remains relatively constant at TP concentrations less than this threshold, but decreases at higher TP concentrations. The location of this potential threshold can be evaluated using discontinuous linear regression (section 4.3).

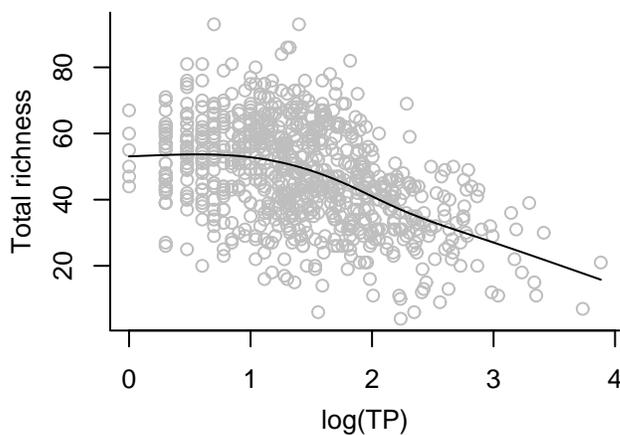


Figure 6. Scatter plot of $\log(\text{TP})$ vs. total richness in the EMAP-West Streams Survey dataset. Non-parametric smooth fit to data shown as solid line. Units are $\mu\text{g/L}$ for $\log(\text{TP})$.

A scatter plot matrix for EMAP-West Streams Survey data (Figure 7) provides a means of viewing simultaneously many combinations of what was demonstrated with a single scatter plot in Figure 6. Combinations of variables from different rows are shown as scatter plots in the intersecting cells and diagonal cells indicate the position of each variable in each graph. Graphs in the bottom half (below the variable names), the units of the variable above are on the x-axis and the variable to the right on the y-axis. In the upper half (above the variable names), the variable to the left is on the y-axis and the variable below on the x-axis. Note that increases in nutrient concentrations (both TP and TN) are associated with decreases in total macroinvertebrate species richness. These nutrient concentrations are also associated with the percentage of substrate sand/fines and with the predicted intensity of grazing activities in the watershed. These observations on relationships among variables can assist in determining an appropriate stressor-response relationship for use in identifying a candidate nutrient criterion.

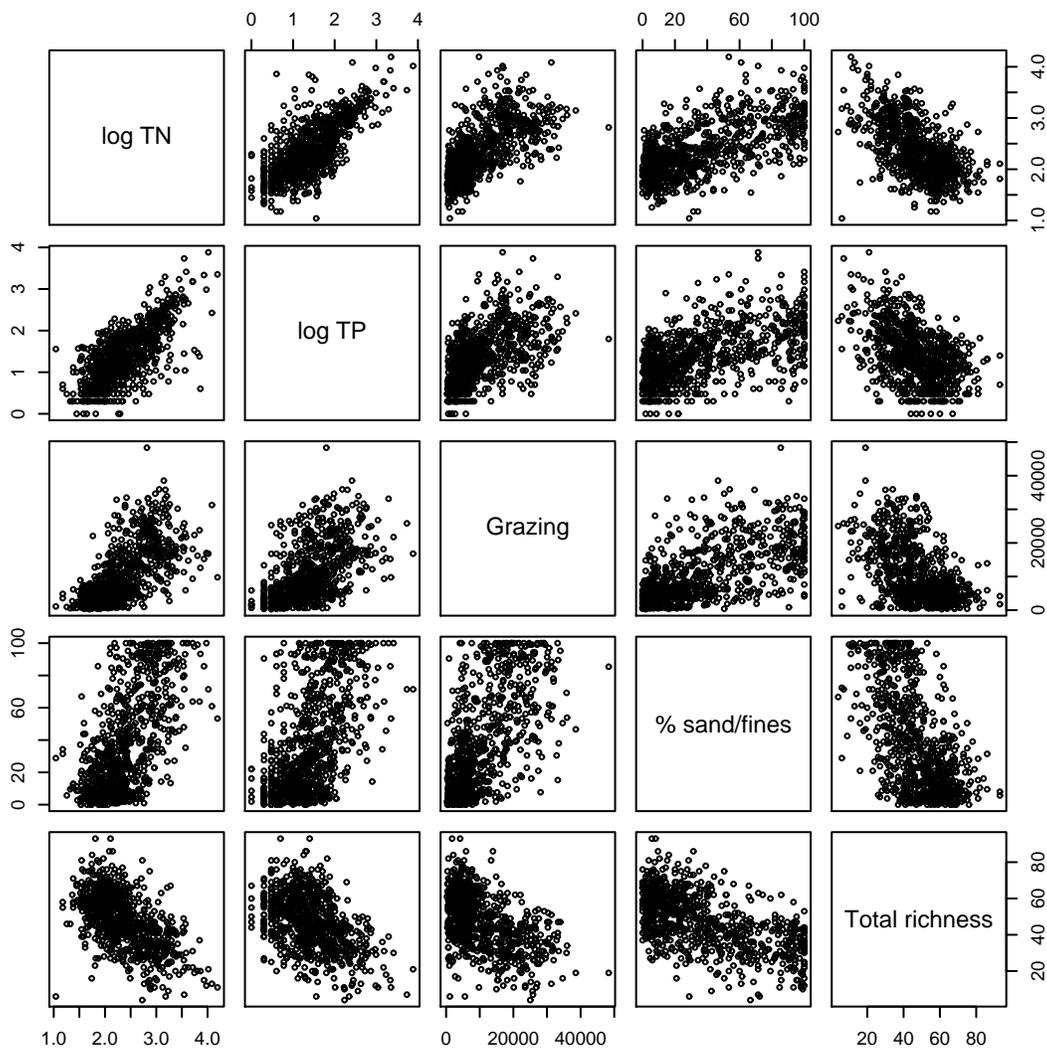


Figure 7. Scatter plot matrix of EMAP-West Streams Survey TN and TP (as log-transformed variables) against measures of animal grazing in the watershed, percent sand/fine substrates and total macroinvertebrate richness. Units are $\mu\text{g/L}$ for TN and TP.

1.4 Maps

Mapping data can provide insights into whether factors vary systematically in space. At a minimum, a map of locations where the data were collected can be produced. A map of the locations of the lake sites that were sampled for the EMAP Northeast Lakes Survey, with symbols sized according to chlorophyll *a* concentration, shows no obvious spatial trends in chlorophyll *a* (Figure 8). The spatial patterns exhibited in maps can provide insight into factors which co-vary with nutrient variables. Although geographic information systems (GIS) are most frequently used to generate maps, exploratory spatial analysis can be done with most graphics and spreadsheet software applications.

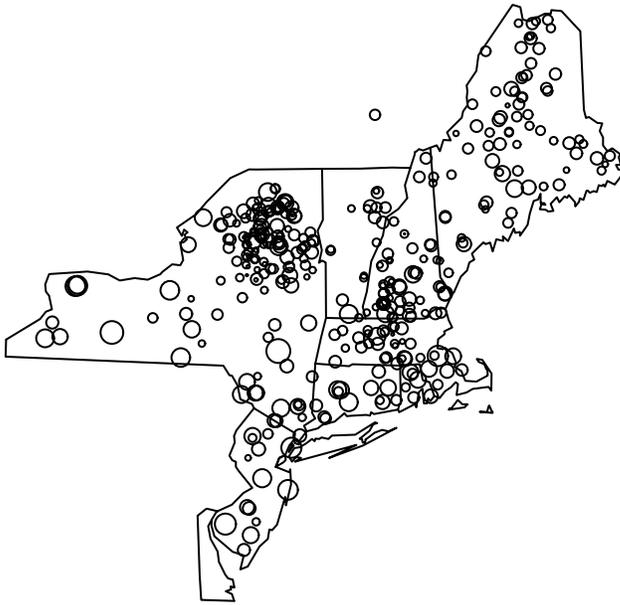


Figure 8. Map of EMAP Northeast Lakes Survey sample sites. Symbol size is proportional to $\log(\text{chl } a)$ concentration.

1.5 Conditional probability

A conditional probability is the probability of an event Y occurring given that some other event X also has occurred. It is denoted as $P(Y|X)$. Thus, a conditional probability describes the probability of observing an event of interest in a subset of samples drawn from the original statistical population. These subsets are defined by conditions when X has occurred, in addition to those used to define the entire statistical population. A conditional probability is calculated as the ratio of the joint probability that Y and X occur simultaneously in a given sample from the original statistical population, $P(Y, X)$, to the probability of X in the original population, $P(X)$. This relationship can be written as follows:

$$P(Y|X) = \frac{P(Y, X)}{P(X)}$$

The possible applications of conditional probability analysis are very broad. In this context, its use is focused on the derivation of nutrient criteria in which conditional probability analysis describes the probability of environmental or ecological impairment (i.e., not meeting the designated use) given that a nutrient concentration is higher than a criterion value. For example, conditional probability analysis can quantify the probability of a benthic community impact given that phosphorus concentrations in the water column exceed $100 \mu\text{g/L}$. A more detailed description of conditional probability analysis is provided in Appendix C.

For use in the exploratory phase, conditional probability analysis (CPA) can screen variables for use in the development of stressor-response relationships. For example, Figure 9 shows CPA plots for the EMAP Northeast Lakes Survey data with chlorophyll *a* as the response (using a threshold of 15 $\mu\text{g/L}$) with TP, TN, dissolved organic carbon (DOC), and turbidity as potential predictors, since each can potentially affect chlorophyll *a*. Each point on a plot displays the probability of lakes exceeding the chlorophyll *a* threshold given that the indicated stressor level is exceeded. For example, chlorophyll *a* exceeded the specified threshold in every lake with $\log(\text{TP})$ concentrations greater than 2 (TP > 100 $\mu\text{g/L}$). This stressor value, at which the observed probability of exceeding the threshold is 100%, could provide an upper bound for candidate nutrient criteria.

A CPA identifies possible stressors to examine for criteria derivation. Additional information or analysis can then be considered to evaluate whether an appropriate causal relationship exists or can be supported.

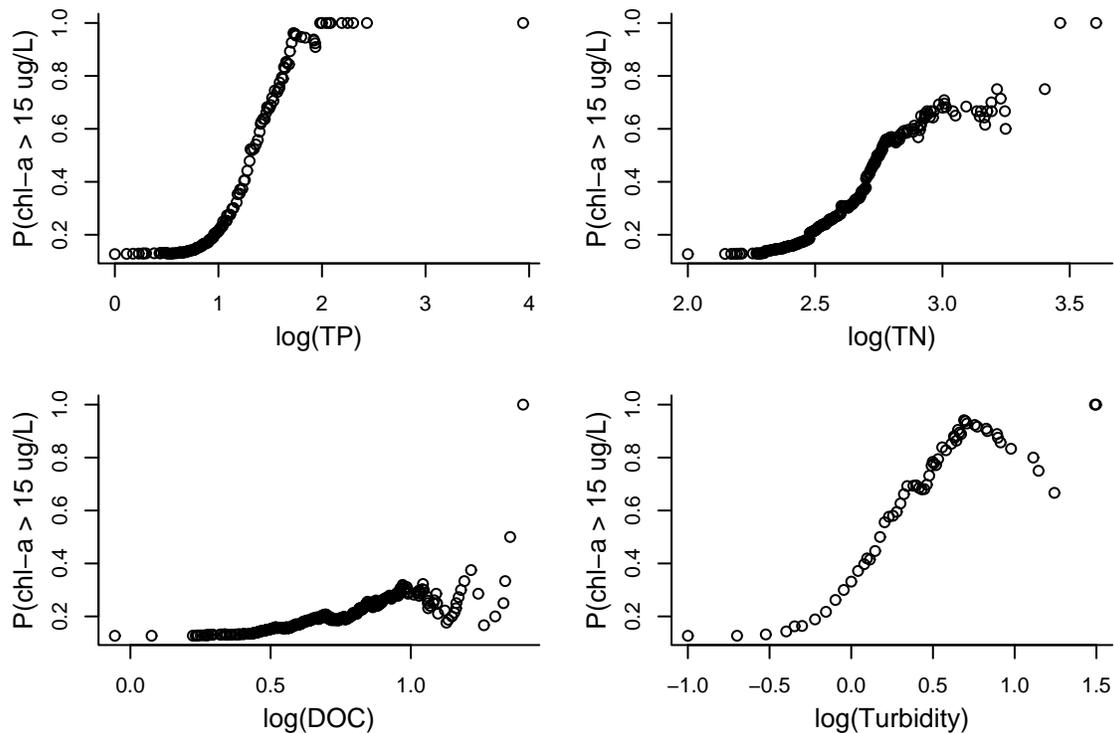


Figure 9. Examples of exploratory data analysis plots using conditional probability analysis (CPA) of EMAP Northeast Lakes Survey data for chlorophyll *a* as a response variable (threshold at 15 $\mu\text{g/L}$) and potential stressor variables of $\log(\text{TP})$, $\log(\text{TN})$, dissolved organic carbon (DOC), and turbidity. Units are $\mu\text{g/L}$ for TN, TP, and DOC and NTU for turbidity.

1.6 Classifications

During data exploration the distributions of selected response and nutrient variables should be examined across different classes (e.g., ecoregions) to determine whether

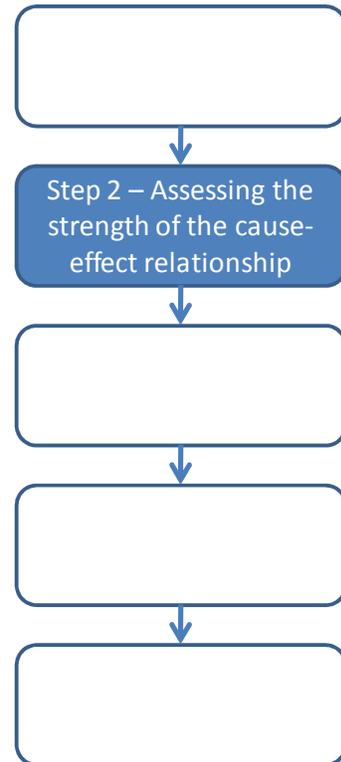
values of these variables differ systematically (e.g., Figure 3, USEPA 2000a). Stressor-response analyses are best conducted for data collected from reasonably similar waterbody types, and classifying data can help ensure this type of homogeneity.

2 Assessing the strength of the cause-effect relationship

Excess nutrients have been shown to alter the structure and function of aquatic ecosystems (Allan 1995, Kalff 2001, Wetzel 2001). These ecological changes may be directly detected by some of the indicators used to determine whether a waterbody supports its designated use or they may be indirectly related to these indicators. Hence, when stressor-response relationships are used to establish nutrient criteria, it is important to assess the degree to which changes in nutrient concentration are likely to cause changes in the chosen response variable (Step 2, Figure 1).

Most stressor-response analyses of nutrients and response variables are based on empirical data collected in the field. Relationships estimated from these observational data can be confounded by other factors (e.g., bedded sediment) that co-vary with the nutrient concentrations, and so, assessing the strength of the cause-effect relationship can be difficult. Laboratory studies provide much stronger causal inferences. However, the applicability of information extracted from these studies and the conditions tested in the experiment need to be evaluated for comparability and relevance to the particular region for which criteria are being derived.

A number of approaches for assessing the strength of a cause-effect relationship have been proposed. For example, a detailed approach for identifying the likely causes of impairment at a single site is available (<http://www.epa.gov/caddis>), and many of these principles can be applied here where the focus is on stressor-response relationships across a population of sites. Methods for assessing causal relationships related to human health have also been used extensively (e.g., Hill 1965, Rubin 1974, and Greenland 2000). In the present context, to assess the strength of the causal relationship between chosen nutrient and response variables, conceptual models linking changes in nutrient concentrations to changes in response variables should first be considered, and existing literature documenting investigations of similar types of responses should be surveyed. Additional alternate methods include statistical analyses designed to evaluate causal effects and experiments.



2.1 Conceptual models

In the context of this report, a conceptual model is a description of the predicted relationship between nutrient stressors and environmental responses. Conceptual models can include both a visual representation (Figure 10) and a written description of the predicted relationships. Conceptual models may be as simple as a two-box model for a stressor and a response, or may be as complex as the many interactions included in some mechanistic models. The conceptual model should identify *a priori* expectations. These expectations may consist of simple positive or negative associations between different variables, or may actually include the strength of the presumed interaction. Many conceptual models linking excess nutrients to ecological changes are readily available in the published literature and on the Internet (e.g., <http://cfpub.epa.gov/caddis/icm/ICM.htm>).

Conceptual models provide valuable tools for communicating both existing knowledge and assumptions regarding the effects of nutrients. Furthermore, conceptual models can help identify other factors that may confound predicted relationships between nutrients and response variables. When observational data are available quantifying different variables included in the conceptual model, it may be possible to test whether observations are consistent with the hypothesized relationships (see Section 3.1.3).

2.2 Existing literature

Existing literature can provide important qualitative evidence in support of a cause-effect relationship between a chosen response and nutrients. Existing literature should be evaluated critically to determine the strength of the causal relationships documented in each study. Randomized controlled studies that document a biological response to increased nutrient concentrations and meta-analyses of these types of studies often provide stronger evidence of a causal relationship than simple analyses of observational data. For example, many manipulative studies of the effects of nutrients on macroinvertebrate grazer taxa have identified reasonably consistent responses (see Feminella and Hawkins 1995), whereas analyses of associations between nutrient concentration and total species richness in observational data have found both increasing (e.g., Heino et al. 2003) and decreasing relationships (e.g., Wang et al. 2007). Study designs used to collect data and the statistical analyses that are applied to the data vary among studies and should be evaluated to determine their relevance and applicability for the context under consideration.

Existing literature may not provide information for precisely the same response variable as has been selected for analysis, but information that considers closely related variables can be useful. In many States and Tribes, macroinvertebrate indices of biological condition (e.g., index of biological integrity or IBI) are used to assess attainment of aquatic life use, and these indices would be natural choices for response variables. Where existing literature is not available or relevant to the particular

multimetric index under evaluation, studies that have focused on individual metrics can also be used and relied upon to provide useful information. For example, as discussed previously, a number of manipulative studies have established strong relationships between nutrient concentrations and macroinvertebrate grazer taxa (Feminella and Hawkins 1995, Hillebrand 2002). Thus, if grazer taxa are included as component metrics in an IBI, these studies could provide supporting evidence of a stressor-response relationship.

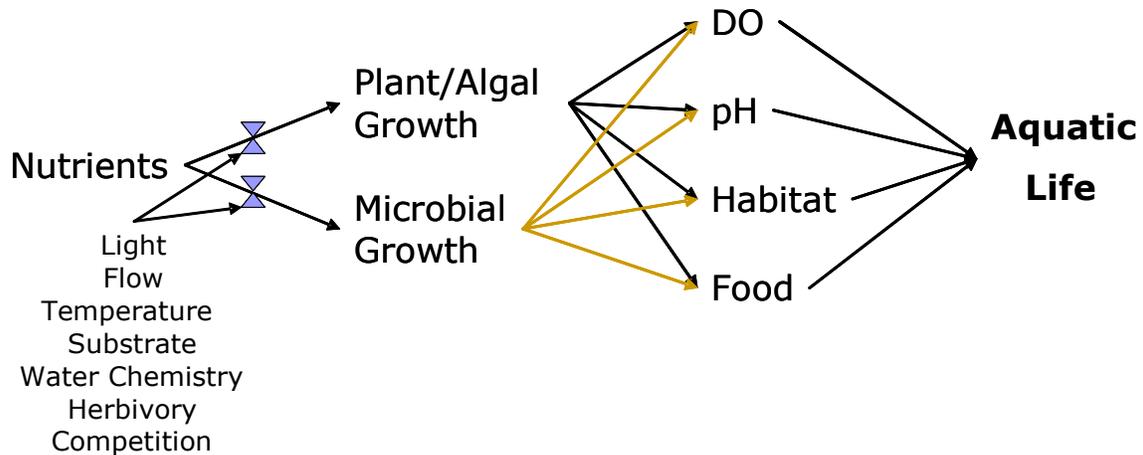


Figure 10. Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts. Nutrients enrich both plant/algal as well as microbial assemblages, which lead to changes in the physical/chemical habitat and food quality of streams. These effects directly impact the insect and fish assemblages. The effects of nutrients are influenced by a number of other confounding factors as well, such as light, flow, and temperature.

2.3 Alternate methods for assessing causal effects

Supporting literature can provide qualitative evidence that the chosen response variable responds causally to changes in nutrient concentrations; however, more involved statistical models, such as structural equation models and propensity score analysis, can potentially provide quantitative support as well. Structural equation models (SEM) are statistical models that test whether hypothesized relationships among variables (e.g., as displayed in a conceptual model) are consistent with the observed covariance structure among those variables (Shiple 2000). In the context of estimating nutrient stressor-response relationships, SEM can provide quantitative support for the predicted relationships described in the conceptual model. For example, the model described in Figure 10 could be analyzed with SEM, testing whether the hypothesized relationships between changes in nutrient concentrations, intermediate variables, and the response variable are consistent with the observed covariance structure. SEMs can provide strong evidence for rejecting or affirming hypothesized pathways and causal linkages (Shiple 2000). They have been applied for analyzing nutrient effects in freshwaters (e.g., Riseng et al. 2004).

Propensity score analysis was developed specifically to more reliably estimate stressor-response relationships from observational data (Rosenbaum 2002). Because many different factors can potentially co-vary with nutrient concentrations, an estimate of the causal effects of nutrients should also consider the effects of these other factors on the response variable. One relatively simple approach for controlling for other factors is to identify samples in the dataset that are similar to the sample of interest with regard to covariate distributions, but differ with regard to the factor of interest. For example, one can identify groups of streams that differ in their nutrient concentrations, but are similar with respect to most other observed environmental factors. If only a single factor (e.g., bedded sediment) co-varied with the factor of interest, one could stratify the dataset by this one factor, splitting the dataset into groups with similar values of bedded sediment. Stratifying by different factors becomes more difficult as the number of factors increases, and propensity score analysis was designed to mitigate this constraint. Propensity scores (Rosenbaum and Rubin 1983, Rosenbaum 2002, Imai and Van Dyk 2004) summarize the contributions of all known covariates as a single variable (the propensity score). As a result, stratifying by propensity score splits the dataset into groups in which all covariate distributions are similar. Then, the causal effects of nutrients can be more confidently estimated within each group because the distributions of other covariates are similar.

Propensity score analysis provides a robust approach for controlling the confounding effects of other factors. Furthermore, propensity score analysis can potentially be directly applied to a field dataset collected from the region of interest, using the selected response variable, and provide quantitative support for the existence of a stressor-response relationship. Propensity score analysis controls for factors for which data are available, and is ideal for large datasets where many different measured variables are available. More detailed information on propensity scores and an example application are provided in Appendix B.

Experimental validation of causal relationships between selected nutrient and response variables can also be considered. This approach involves using specific experimental designs (e.g., controlled experiments) employing in situ manipulations, mesocosms, and/or laboratory experiments. Such approaches have long been used to test nutrient effects in aquatic systems (e.g., Schindler 1974, Elwood et al. 1981, Schindler et al. 1987, Peterson et al. 1993, Rosemond et al. 1993, and Cross et al. 2006).

3 Analyzing data

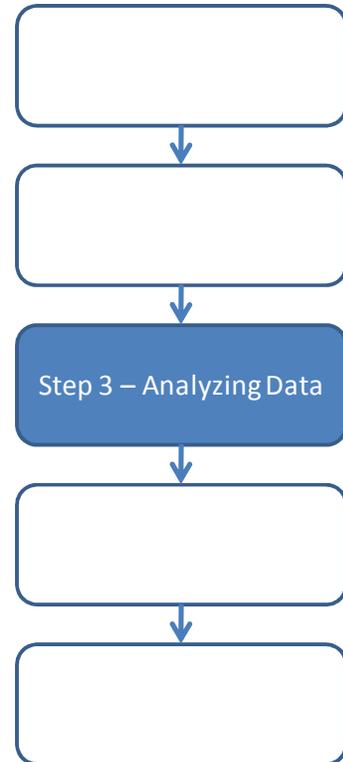
In Step 3 of the process (Figure 1), data are analyzed to estimate stressor-response relationships. Analyses are presented in two main subsections that provide different approaches for deriving criteria: (1) translating existing response thresholds into candidate criteria using estimated stressor-response relationships and (2) identifying thresholds from inherent characteristics of the stressor-response relationships. Each subsection begins with an overview of the process by which an estimated stressor-

response relationship is used to derive a nutrient criterion value. Then, statistical methods are described, the data requirements for each are reviewed, and example applications relevant to nutrient criteria derivation are presented.

3.1 Translating a response threshold to a candidate criterion: single explanatory variable

In many cases, a threshold for the selected response variable is available that defines values of the response variable where designated uses are supported. For chemical acute water quality criteria, the USEPA has defined this threshold as the lower 5th percentile of the distribution of applicable acute values to represent a low overall effect level to species in the broader ecosystem. This threshold was judged to be protective of all species (Stephan et al. 1985). A comparable approach is not applicable to deriving water quality criteria for nutrients because adverse effects to the designated use of a waterbody occur at concentrations of nutrients below the level which is shown to be toxic to organisms. Alternative approaches for establishing thresholds for nutrients are available, though. For example, a protective level may be pre-determined if criteria already exist in State standards to protect the designated use (e.g., biological criteria). Also, expert opinion regarding protective levels of variables can be methodically collected (Reckhow et al. 2005), and surveys can identify conditions that support user expectations for different waterbodies (Heiskary and Walker 1988).

Some response variables may be indirectly associated with a designated use, and therefore, specifying a direct response threshold may be more difficult. One possible analytical approach is to model the relationship between the selected response variable and a variable for which a response threshold is available. For example, data exploration may suggest that one characteristic of the aquatic community responds strongly to increased nutrients (e.g., oligochaete abundance), but this metric does not directly measure attainment of any designated uses. One could potentially model the relationship between oligochaete abundance and a multimetric index for which threshold values are available. This auxiliary model would then provide a threshold for use in establishing a candidate nutrient criterion from the relationship between oligochaete abundance and nutrient concentrations. Applying this approach requires considerably more statistical modeling, and evaluating the uncertainties between the two models would require careful consideration. In general, focusing attention on response variables that can be more clearly linked to designated uses will result in simpler analyses.



Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

When a threshold value is available for the selected response variable, then an estimated stressor-response relationship can be used to “translate” this response threshold into a comparable nutrient criterion (Figure 11). The horizontal red line in Figure 11 represents the point at which designated uses are supported, and the intersection between the solid black line and the horizontal red line indicates the nutrient concentration at which, on average, predicted values of the response variable are equivalent to the threshold value. This nutrient concentration then provides a possible criterion (label A, in Figure 11). For example, in lakes, a criterion for water clarity, expressed in terms of Secchi depth may already exist, and this criterion could be used in conjunction with an empirical relationship between TP and Secchi depth to estimate a protective phosphorus concentration.

In all models of empirical data, variability in the stressor and response measurements, and uncertainty in the model structure influence prediction uncertainty. That is, for any specified nutrient concentration, predictions are more accurately characterized as a range of possible values of the response variables, given any particular nutrient concentration (dashed lines in Figure 11). In some situations, this uncertainty can be incorporated into the criterion determination. In the example shown in Figure 11, the lower prediction interval represents a boundary above which approximately 95% of observations are located. Then, the intersection between the lower prediction interval and the red horizontal line provides a nutrient concentration (label B, Figure 11) at which the value of the response variable at approximately 95% of sites will be at least the value of the specified response threshold.

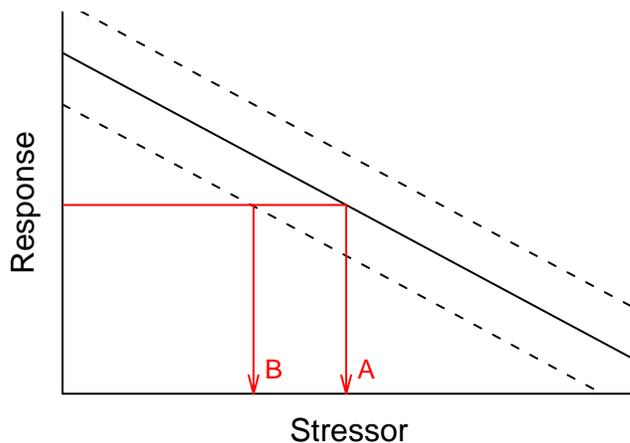


Figure 11. Plot showing example of stressor-response relationship, where the response is a direct measure of designated use or can be easily linked to a designated use measure. The solid line is the mean estimated relationship. The dashed lines are the upper and lower 95% prediction intervals. The red horizontal line indicates the response variable threshold, and the vertical red lines indicate two possible criteria (A and B).

Note that although the relationship shown in Figure 11 is linear, it does not preclude applying this approach to relationships of any form (e.g., non linear relationships as

shown in Figure 6). In this section, methods are presented for performing the analytical approach described here.

3.1.1 *Simple linear regression (SLR)*

Description and Data Requirements

Simple linear regression is a method for testing the relationship of a dependent variable (y-axis) with some independent variable (x-axis) and for developing a predictive relationship between the two. SLR is a well-established method for conducting such analyses and is described in the existing nutrient criteria guidance (USEPA 2000a). The result of a simple linear regression is a line that describes the relationship between the two variables and the coefficients for that line. Simple linear regression uses a least squares function to calculate the coefficients for the line.

SLR assumes that the XY pairs of points are numeric data that represent independent samples from some underlying population. For example, a typical dataset for use in deriving nutrient criteria would consist of independent samples from different streams in a particular State or region. The range of conditions represented by the data should span the range of conditions over which one wishes to evaluate relationships and/or make predictions. Predictions extrapolated outside this range are more uncertain. SLR assumes the relationship is linear and that the error term of the model (unexplained variance) is normally distributed and exhibits constant variance. Various transformations exist for one or both variables if one expects the relationship to be non-linear. Frequently used transformations with water quality data include the common and natural logarithms applied to data that are strongly skewed with many relatively small values and a few very large values, and arc-sine square-root transformations, that are applied to data measuring proportions or percentages. Transformations can make interpretation more complex and may require back-transformation. In some cases, one can analyze non-linear responses using non-linear regression techniques.

The results of a SLR are commonly evaluated using several diagnostic statistics, which can quantify characteristics of the relationship estimated between the nutrient variable and response variable and help assess the appropriateness of regression model assumptions. Frequently calculated statistics include the statistical significance of the model coefficients, which tests whether the estimated model slope and intercept are significantly different from 0. The statistical significance of the slope of the estimated relationship between the nutrient variable and the response variable provides some indication of whether the relationship exists beyond what one would expect from chance alone. The coefficient of determination (R^2) is a measure of the proportion of variance in the response that is explained by the regression model. Plotting the difference between each observed value and its mean predicted value (i.e., the residuals) against the independent variable provides a qualitative means of assessing

Empirical Approaches for Nutrient Criteria Derivation SAB Review Draft

the assumption that residual variance is constant across the range of modeled conditions. Other diagnostic tests include tests for leverage or influential points.

When used for deriving nutrient criterion values, SLR is applied in a predictive mode, so it is important to verify the underlying assumptions of the method using the diagnostic tests described above. Model predictions can also be validated with independent data or with bootstrapping or cross-validation techniques [see Section 5 (Step 4)].

SLR has been applied frequently in environmental analysis and, specifically, in nutrient criteria analysis. The *Nutrient Criteria Technical Manual for Rivers and Streams* (USEPA 2000a) includes the results of SLR of periphyton chlorophyll *a* versus TP for over 200 streams around the world (Dodds et al. 1997) and references the often used predictive equation for suspended chlorophyll versus TP from 292 temperate streams (Van Nieuwenhuysse and Jones 1996).

Example Applications

Three examples are provided of applying SLR for criteria derivation. In the first example, the relationship of TP concentration to the concentration of phytoplanktonic chlorophyll *a* in northeastern lakes as part of the EMAP Northeast Lake Survey was tested (Figure 12). Matched pairs of data from lakes were used and represent grab samples of nutrients taken at the time of chlorophyll sampling (see Appendix A for data description). In this application, the question was whether chlorophyll *a* concentrations were significantly associated with TP concentration. Initial correlation analysis suggested the two variables were related and SLR was used to test the strength of the relationship and to generate a predictive model from the data. Chlorophyll *a* did increase with TP, but both chlorophyll *a* and TP concentrations varied over a wide range and their distributions were heavily skewed (i.e., appear to be non-linear). By log-transforming TP and chlorophyll, the values better fit a linear model because they were more uniformly distributed across the x-axis, and the relationship was approximated with more confidence.

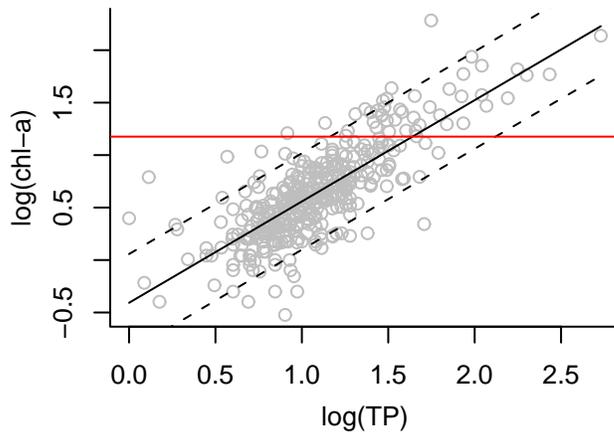


Figure 12. Log(TP) versus log(chl *a*) using EMAP Northeast Lakes Survey data. Solid line: mean regression relationship, dashed lines: 90% prediction intervals. Red horizontal line indicates chlorophyll *a* = 15 µg/L. Units are in µg/L. Regression Equation: $\text{Log}(\text{chl } a) = -0.41 + 0.97[\text{log}(\text{TP})]$; $R^2=0.61$, $p<0.001$.

The SLR results show that in this model, every order of magnitude increase in TP was associated with an order of magnitude increase in chlorophyll *a*. Log TP explains 61% of the variation in log chlorophyll. Other factors, including other nutrients, influence chlorophyll in these lakes and it may also be that factors co-varying with TP explain a portion of the 61%. Prediction intervals (shown as dashed lines in Figure 12) provide an estimate of the range of possible chlorophyll *a* values predicted for a given log(TP) value.

Chlorophyll is used as a predictor of trophic state in lakes (Carlson 1977) and a value of 15 µg/L has been used as a cutoff between mesotrophic and eutrophic conditions (Vollenweider and Kerekes 1980). If one wished to protect the designated use against eutrophy based on this threshold value and data shown in Figure 12, then, the regression equation could be used to predict a mean TP value associated with that chlorophyll threshold of 15 µg/L (red horizontal line of Figure 12). Using the equation defining the regression line in Figure 12 and remembering to convert the resulting criterion from a log-transformed value back to its original units, a mean chlorophyll concentration of 15 µg/L would be predicted to occur on average at a TP concentration of $\text{log}(\text{TP}) = 1.6$ (TP = 40 µg/L).

Prediction intervals provide estimates of the range of possible chlorophyll *a* values that might be observed for a given value of log(TP), and different prediction intervals can be used to define criteria with different degrees of “protectiveness”. For example, lakes below a TP concentration of 40 µg/L have less than a 50% chance of exceeding the chlorophyll *a* threshold of 15 µg/L. Similarly, the upper prediction interval in Figure 12 indicates that lakes below a log(TP) concentration of 1.2 (TP = 16 µg/L), have less than a 10% chance of exceeding the chlorophyll threshold. Other criteria corresponding to different prediction intervals can be calculated, but note that estimates of percentages of lakes maintaining the desired biological threshold apply only to lakes with TP

Empirical Approaches for Nutrient Criteria Derivation SAB Review Draft

concentrations that are equivalent or less than the specified criterion. Estimates of the effects of different candidate criterion values on all lakes in the study area require further calculations [see Section 6 (step 5)].

The second example illustrates the relationship between nutrients and macroinvertebrate metrics. This relationship represents two variables separated by a greater distance along the causal pathway because nutrients are generally not directly toxic to macroinvertebrates except at very high concentrations, but may impact macroinvertebrates negatively at lower concentrations via enrichment pathways (Figure 10). Many States use macroinvertebrate metrics, such as tolerant species richness or total species richness, for assessing biological condition (USEPA 2002), and macroinvertebrate metrics and indices formed from combinations of these metrics often provide a direct measurement of the degree to which a waterbody supports aquatic life use (Kerans and Karr 1994, Barbour et al. 1996, Klemm et al. 2003).

Here, total species richness was plotted against log-transformed TN (Figure 13) using data from the EMAP-West Stream Survey Xeric region stream dataset (herein referred to as EMAP West Xeric region streams), collected by the USEPA EMAP-West Streams Survey (Appendix A). A threshold value of total stream macroinvertebrate species richness typically would be defined as part of State water quality standard as the point at which aquatic life use is supported, and has been assumed to be 40 for illustrative purposes (red horizontal line on Figure 13). Total species richness decreased with increased TN, and the mean regression prediction intersects the assumed biological threshold at $\log \text{TN} = 2.5$ ($320 \mu\text{g/L}$). This value defines a candidate criterion below which streams would have a greater than 50% chance of maintaining the desired biological response threshold of 40 taxa.

Prediction intervals around the mean relationship are larger than in the previous example, and the lower 10% prediction interval does not intersect the biological threshold line. A criteria could be easily estimated, however, based on the mean regression for where a greater than average chance of maintaining the desired biological response threshold of 40 taxa exists, and for several percentiles below this but not for 10%. Including additional variables in the regression model may reduce the residual variance and alter the resulting criterion values.

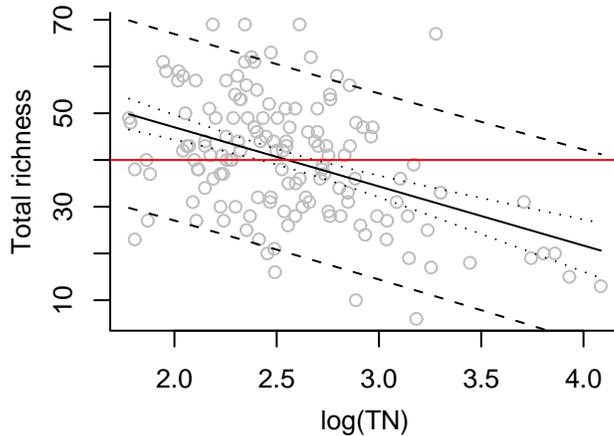


Figure 13. Log(TN) versus total species richness in EMAP-West Xeric region streams. Solid line: mean regression relationship, dashed lines: 90% prediction intervals. Red horizontal line indicates total richness = 40. Units are $\mu\text{g/L}$ for log(TN). Regression equation: Total Richness = $72 - 13[\text{Log(TN)}]$, $R^2=0.19$, $p<0.001$.

A third example was developed using data on nutrients and stream benthic macroinvertebrates from the Maryland Biological Stream Survey (MBSS) program (<http://www.dnr.state.md.us/streams/mbss/>). EPT species richness calculated from MBSS macroinvertebrate samples from the Piedmont ecoregion was combined with total phosphorus data from those same streams and used to explore the response of EPT richness to increasing nutrient concentrations. Significant portions of the Piedmont ecoregion are urbanized. Urbanization generates other stressors that impact macroinvertebrates in addition to nutrients. Therefore, Piedmont sites with high amounts of urbanization were removed to reduce covariance and focus more directly on nutrient effects. EPT richness declined significantly with increasing total phosphorus in these Piedmont streams, although TP explained very little of the decline in EPT richness ($r^2 = 0.05$, Figure 14). The MBSS scores EPT richness based on expectations for least disturbed streams and the midpoint of the distribution of EPT scores from least disturbed streams to most disturbed Piedmont streams is an EPT richness of 8. From the figure, an EPT richness of 8 is associated, on average, with a TP concentration of 180 $\mu\text{g/L}$ TP. As with the example above, a concentration of TP could be estimated below which sites would have a greater than 50% chance of meeting the EPT threshold and greater percent likelihoods above 50% could be estimated, but in this case not the 90% chance because the lower decile (hatched line) does not intersect the threshold.

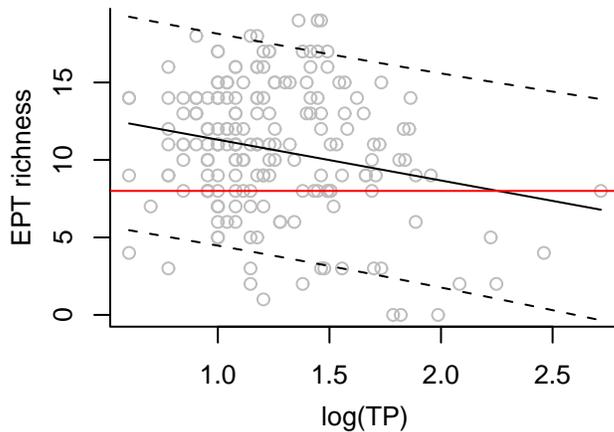


Figure 14. Log(TP) versus EPT species richness using the least urban MBSS Piedmont ecoregion stream dataset. Solid line: mean regression relationship, dashed lines: 90% prediction intervals. Red horizontal line indicates EPT richness = 8. Units are $\mu\text{g/L}$ for log(TP). Regression equation: $\text{EPT Richness} = 13.9 - 2.63[\text{Log(TP)}]$; $R^2=0.05$, $p<0.005$.

3.1.2 Quantile regression (QR)

Description and Data Requirements

Regressions on ecological data often exhibit unequal residual variance due to missing explanatory variables and interactions among the factors affecting biological conditions. When residual variance is unequal, the range of the scatter of samples about the mean regression line differs for different parts of the line, and more than a single slope (rate of change) may describe the relationship between the response and nutrient variable. Quantile regression (QR) is a statistical method that can quantify these relationships by estimating linear relationships for different quantiles of data (Koenker and Bassett 1978, Koenker and Hallock 2001, Cade and Noon 2003, Koenker 2005, Brenden et al. 2008). Just as classical linear regression methods based on minimizing sums of squared residuals enable one to estimate models for mean functions; QR methods offer a mechanism for estimating models for the median value and the full range of other quantiles.

QR can be useful for informing criterion selection for responses that do not satisfy the assumptions of SLR. As described above, when the magnitude of residual variance differs for different predictor values, SLR estimates of prediction intervals can be incorrect. QR (e.g., for 5th and 95th percentiles) provides a direct means of estimating these prediction intervals, and comparison of QR estimates of these prediction intervals with SLR estimates provides one means of testing SLR assumptions. QR also can directly identify a relationship between a stressor and response when one suspects that a particular factor sets an upper limit to the value of the response variable (Cade and Noon 2003), and such applications have been used to set criteria for conductivity (FDEP 2008) and sediment (Bryce et al. 2008). Other variations of QR exist that combine QR

with, for example, split-point regression and extend the approach for identifying thresholds (Brenden et al. 2008, Brenden and Bence 2008).

Example Applications

QR was applied to the EMAP Northeast Lakes Survey chlorophyll a data (Figure 15). The 95th prediction intervals from quantile regression were similar to that predicted by SLR, suggesting the regression assumptions for SLR were appropriate. The resulting candidate TP criterion based on the median value was the same as predicted by mean SLR (40 $\mu\text{g/L}$).

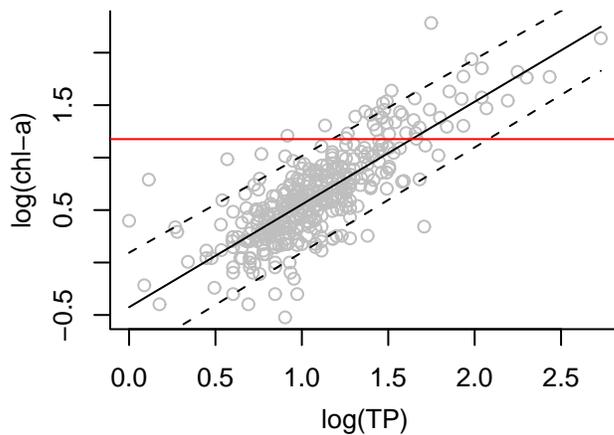


Figure 15. Log(TP) vs. log(chl a) for EMAP Northeast Lakes Survey. Dashed lines are the 5th and 95th percentile estimated by quantile regression. Solid line is the 50th percentile. Red line indicates where chlorophyll a = 15 $\mu\text{g/L}$. Units are $\mu\text{g/L}$.

Quantiles estimated for the relationship between total species richness and log(TN) in EMAP-West Xeric region streams exhibited more variability than observed in the northeastern lakes example (Figure 16). Much of this variability can likely be attributed to uncertainties associated with estimating extreme quantiles (e.g., 5th and 95th quantiles) from relatively small sample sizes. In the present case, the 143 samples that were available likely were insufficient for accurately estimating the 5th and 95th quantiles. This variability is important information, and some of the differences in slope observed across different quantiles suggest that regression assumptions, such as constant variance, are not supported by these data. The 50th percentile line crosses the biological response threshold at approximately log(TN) = 2.4 (TN = 250 $\mu\text{g/L}$), a value that is lower than that given by the mean SLR relationship and below this concentration, there is a greater than 50% chance of meeting the total richness threshold. Greater percent likelihoods above 50% could be estimated, but in this case not the 90% chance because the lower decile (hatched line) does not intersect the threshold.

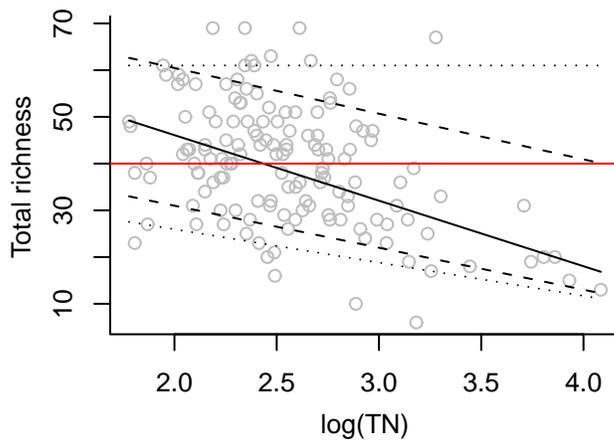


Figure 16. Log(TN) vs. total species richness for EMAP-West Xeric region streams. Dashed lines indicate 10th and 90th quantiles estimated with quantile regression, dotted lines indicate the 5th and 95th quantiles, and the solid line indicates the 50th quantile. Units are $\mu\text{g/L}$ for log(TN). Red line indicates where total species richness = 40.

A second example was developed using data on nutrients and stream benthic diatoms from the USEPA EMAP Mid-Atlantic Integrated Assessment (MAIA) and United States Geological Survey (USGS) National Water Quality Assessment (NAWQA) program. Diatom data collected by the two programs were combined and a trophic state index (TSI) based on the van Dam trophic state index (van Dam et al. 1994) was calculated using the diatom information (J. Stevenson, pers. comm.). Values of this index indicate increasing trophic state, much as the lake TSI does (Carlson 1977). A value of 4.5 indicates a shift from mesotrophic to eutrophic conditions. TSI values increased significantly with TP concentration, and the TP concentration associated with a TSI value of 4.5 using the median regression was $\log(\text{TP}) = -1.55$ ($\text{TP} = 0.028 \text{ mg/L}$) (Figure 17). As with the EMAP-West Xeric example above, a concentration of TP could be calculated that would result in a greater than 50% chance of sites meeting the response threshold. As above, percent likelihoods above 50% could be estimated, but not the 90% chance because the upper decile (hatched line) does not intersect the threshold.

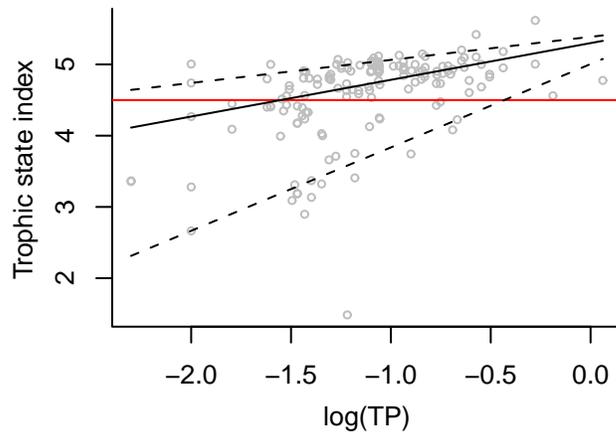


Figure 17. Log(TP) vs. benthic diatom TSI for the EMAP-MAHA/USGS-NAWQA regional dataset. Dashed lines indicate 10th and 90th percentiles estimated with quantile regression, and the solid line indicates the 50th percentile. Units are mg/L for log(TP). Red line indicates where TSI = 4.5.

3.1.3 Logistic regression

Description and data requirements

Logistic regression predicts the probability of an event occurring as a function of different explanatory variables. Thus, the dependent variable in a logistic regression takes only two possible values: either the event occurred in a sample, or it did not. In the context of criterion development, the event would be an observation of an impairment (e.g., chlorophyll *a* concentration > 15 $\mu\text{g/L}$). In most other respects, logistic regression is very similar to SLR. As with SLR, the model can be fit to observed data using maximum likelihood estimation, and statistical tests are used to determine whether model coefficients differ significantly from zero.

Example application

Logistic regression was applied to the stream macroinvertebrate example considered previously for SLR. Here, though, the assumption that total richness values below 40 indicate impairment was applied before fitting the model, by assigning a value of 1 to samples in which the threshold was exceeded and 0 to all other samples. The modeled probability of impairment is shown in Figure 18.

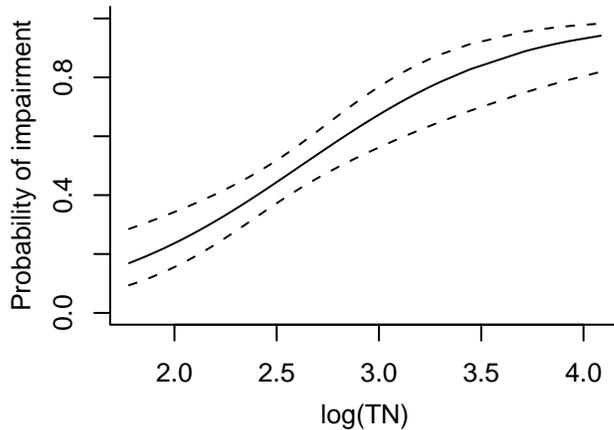


Figure 18. Modeled probability of impairment (i.e., total richness < 40) as a function of log TN for EMAP-West Xeric region streams. Units are $\mu\text{g/L}$ for $\log(\text{TN})$. Solid line: mean relationship, dashed lines: 90% confidence intervals for mean relationship.

The probability of impairment (i.e., total richness < 40 taxa) increased as TN concentration increased. A candidate criterion value could be identified from logistic regression results by selecting an acceptable probability of impairment. For example, the intersection between a 50% probability of impairment and the regression curve specifies a candidate criterion value of approximately $\log(\text{TN}) = 2.6$ ($\text{TN} = 400 \mu\text{g/L}$). Uncertainty in the location of the mean relationship could be incorporated by examining the intersection between different confidence intervals and the desired target benchmark. This value differs from SLR-based and quantile regression-based candidate criteria values that are predicted to maintain 50% of sites at or above the desired biological threshold. These differences across modeling approaches recommend that certain regression assumptions be confirmed for consistency.

In this application, continuous data (i.e., total species richness) has been transformed to binary data (i.e., impaired vs. not impaired), and inherent to this data transformation is a loss of information. Therefore, applying logistic regression as illustrated here may not fully take advantage of the information in the dataset. Logistic regression is best applied to response variables that are originally expressed in terms of binary alternatives (e.g., presence or absence of particular valued species).

3.1.4 Application of bivariate models

The methods described thus far model the effects of a single stressor variable on a single response variable. This section reviews factors related to the use of bivariate stressor-response models for deriving nutrient criteria. These issues should be considered when interpreting the conclusions from bivariate analyses and in determining how the results should be incorporated into the criteria derivation process.

Bivariate analyses can provide straightforward and simple visual and statistical measures of specific responses to specific nutrient variables. Most are fairly easy to interpret and to explain to audiences without substantial scientific or analytical experience. They generally provide valuable interpretations that are consistent with what can be visually observed in the data. They can deduce valid and useful predictive relationships that can be effectively applied to manage environmental results. Lastly, these analyses are also relatively easy to produce and are available with most desktop statistical and spreadsheet software packages.

An important aspect of bivariate models is that they may not account for all the variables known to affect the response variable and therefore, can oversimplify relationships between nutrient and response variables. An analyst is assuming that the nutrient variable being modeled is the major factor affecting the response and is dominant over the range of interest. Other factors may co-vary with the nutrient variable being modeled, and it may be that the covariates affect the response more strongly than the selected nutrient variable. Methods for better understanding the relationships of such co-variates are addressed in 4.2 and subsequent sections.

3.2 Translating a response threshold to a candidate criterion: multiple explanatory variables

In this section, the concepts involved with inferring a candidate criterion based on relationships between a response with a known stressor threshold level and a nutrient concentration are extended to cases where multiple predictor variables are considered, especially ones that may also affect the response. Multiple linear regression models are described here but other analytical options exist including propensity score analysis which was introduced and explained in section 3.1.3 and is further elaborated in Appendix B.

3.2.1 Multiple linear regression (MLR)

Description and Data Requirements

MLR is frequently used in environmental analyses. It is appropriate for attempting to predict responses that may be influenced by multiple factors, for identifying those factors, or for controlling for the effect of certain factors when testing the effect of a stressor. In the context of nutrient criteria derivation, MLR has been applied for classification as well as for identifying candidate criteria (Dodds et al. 2002, 2006, Soranno et al. 2008).

Multiple linear regression is an extension of SLR where the effect of each explanatory variable is modeled as a linear function. In its simplest form, each explanatory variable exerts an effect on the response that is independent of the values of the other variables. Values of the coefficients in MLR models are predicted using the same general

approaches used in SLR, the data requirements are the same, and the diagnostics include most of the same statistics: variance explained, significance of the model and each parameter, residuals inspection, etc. The assumptions for MLR are the same as those for SLR, namely that the error terms are independent, have constant variance, and are normally distributed. Generally, strong collinearity among variables should be evaluated because they may artificially inflate the variance explained by the model, and may affect model fit.

Selecting the appropriate explanatory variables to include in addition to the nutrient variable is a key aspect of using MLR to estimate nutrient criteria, and these selections should be driven based upon an understanding of the linkages between different variables in the studied system (Figure 10). Some automated methods (e.g., stepwise variable selection) are available for selecting predictor variables that best account for observed variability in the response and should be used to help explore possible models, but not necessarily for final model selection (Harrell et al. 1996).

Various techniques can be used to avoid overfitting, in which models apply only to the calibration data and have very poor predictive power outside the calibration data. One rule of thumb regarding the appropriate number of predictors suggests that at most 1 predictor variable is allowed per 10 independent samples (Harrell et al. 1996). Validating models with independent data is also valuable and strongly encouraged [see Section 5 (Step 4)]. Statistical criteria are also available to select the optimal model predictors (e.g., Akaike Information Criterion, Adjusted R^2). A variety of additional analyses are available for building MLR models, including methods for identifying outliers and leverage points, alternative model selection procedures, and additional diagnostics.

Recently, more sophisticated multilevel modeling approaches using Bayesian analysis have been used to refine nutrient-chlorophyll dose-response models using multiple predictors (Lamon and Qian 2008). These models combined classificatory predictors based on ecoregion, and sampling methods with nutrient-algal response models to generate a family of stressor-response models to use under different combinations of conditions (ecoregion and water quality methods for chlorophyll and N and P measurement). Similarly, hierarchical partitioning models (Chevan and Sutherland 1991, MacNally 2000, 2002, Morrice et al. 2008) and partial correlation models (King et al. 2005) have also been used to determine the influence of multiple intercorrelated stressor variables on water quality.

Example Applications

In the first example MLR is used to model the effects of TN on total macroinvertebrate species richness in EMAP-West Xeric region streams. Here, as a first illustration of the use of MLR, the effects of substrate composition, quantified as percent sand/fine

sediment (SED), on total richness are modeled in addition to log(TN). The resulting, fitted model can be written as follows:

$$\text{Total richness} = 67 - 7.5 \log(\text{TN}) - 0.18 \text{ SED}$$

This model explains 32% of the variability in total richness, and both predictor variables are significant ($p < 0.001$). The coefficient for TN estimated here is comparable to coefficients estimated with propensity score analysis (Appendix B), and so effects observed in this simple model seem consistent with effects estimated from more involved analyses.

Because only two explanatory variables are used, the model can still be effectively visualized using scatter plots (Figure 19) by plotting the combinations of the explanatory variables as symbols and by superimposing contours that indicate the predicted mean species richness for any combination of explanatory variables. As with previous examples, a response threshold of total richness equal to 40 is used for illustrative purposes only (shown as a red line).

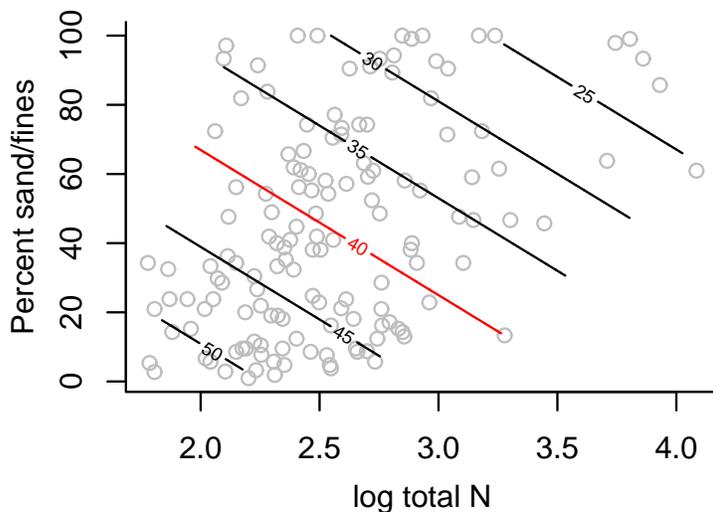


Figure 19. Log(TN) versus percent sand/fines in EMAP-West Xeric region streams. Contours indicate predicted mean values of total richness from multiple linear regression model. Units are $\mu\text{g/L}$ for log total N.

Protective criterion values for TN vary with the value of SED at each site. That is, as SED increases, the value of TN required for maintaining 50% of sites at a minimum total richness of 40 decreases. One possible approach for specifying a criterion using this model is to evaluate the TN criterion that corresponds to different conditions observed for SED. For example, one could use the 75th percentile of SED as a target, assuming one needs to protect against relatively poor sediment conditions. This SED value is equal to 65 % sand/fines. Fixing SED at this value would result in a log(TN) candidate criterion value of 2.05 (112 $\mu\text{g/L}$) to maintain total richness, on average, at 40. The percentile

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

assumed for sediment could be adjusted depending on reasonable management expectations for sediment given natural conditions and rehabilitation potential. For example, one could assume that the criterion should be estimated based on an assumption that no other factors are negatively influencing the response variable. In this case, one would select a relatively low percentile for SED (e.g., the 25th percentile) and calculate a corresponding TN criterion.

A model of the same data was built incorporating an additional natural variable and is given by the equation:

$$\text{Total richness} = 27 - 8.4 \log(\text{TN}) - 0.14 \text{ SED} + 16 \log(\text{PPT}),$$

where PPT= average annual precipitation. Solving again for the relatively worst case scenario of the 75th percentile of sediment and the 25th percentile of annual precipitation (29.5 inches) representing more stressful, dry conditions would yield a target log(TN) of 2.04 (TN = 110 µg/L).

In the next example, the effects of two nutrients on chlorophyll *a* in the EMAP Northeast Lakes Survey dataset were explored. Chlorophyll *a* may be limited by both N and P. In this model, the influence of TN and TP are included simultaneously in predictions of chlorophyll *a*. The same dataset used in the SLR example to show that chlorophyll increased with TN was used for this example.

Both predictors (TN and TP) were statistically significant in MLR model ($p < 0.001$). The final model is given by:

$$\text{Log(Chlorophyll)} = -1.19 + 0.80 \text{ Log(TP)} + 0.37 \text{ Log(TN)} .$$

This model explained 63% of the variance in log(chl *a*) across these lakes, which is a small increase over the percentage of variance explained by SLR. A scatter plot of log(TP) versus log(TN) indicates that the two nutrient concentrations are strongly correlated (Figure 20). The approach described above of fixing one explanatory variable at a specified percentile can be used here as well, but with a caveat. In this case, fixing log(TN) at its 25th percentile value and calculating a criterion for log(TP) would result in a combination of TN and TP concentrations that was not observed in the dataset. This can be used with this understanding, but one could also use the SLR model, which assumes that the effects of both TN and TP can be modeled with TP only and was nearly as effective in explaining observed variability in chlorophyll *a*. The use of SLR in this situation, with known covariates, will reflect the influence of the variable being modeled and all variables that covary with it.

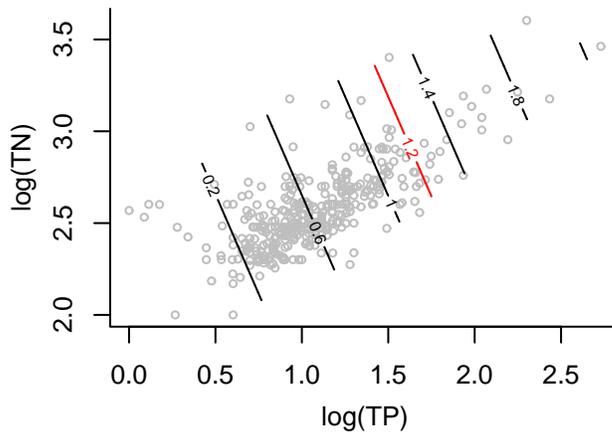


Figure 20. Log(TP) versus log(TN) for EMAP Northeast Lakes Survey sites. Contours indicate predicted mean log(chl a) concentrations. Units are $\mu\text{g/L}$ for log(TP) and log(TN). Red contour indicates assumed biological response threshold of chlorophyll a = 15 $\mu\text{g/L}$.

A multiple linear regression model was also built for the MBSS example for EPT richness and total phosphorus introduced in section 4.1.1. In this instance, epifaunal substrate score (EPISUB), a habitat metric measured by MBSS as part of their stream assessment protocol, was also a significant predictor of EPT richness. Epifaunal substrate score reflects the quality of the faunal substrate and higher scores indicate less fine sediment, more stable substrate, particularly large cobble and/or stable woody debris. As with the EMAP West example above, protective TP values change with the quality of the epifaunal substrate. The multiple linear regression for both predictors was:

$$\text{EPT richness} = 6.9 - 1.3 \log(\text{TP}) + 0.4 (\text{EPISUB}).$$

This model explained 14% of the variance in EPT richness ($p < 0.001$), and each predictor contributed to the model (EPISUB, $p < 0.001$; TP, $p < 0.15$). The significance of TP declined as a predictor when combined with epifaunal substrate in the MLR model relative to the SLR model.

As epifaunal substrate condition declines, the value of TP required for maintaining 50% of the sites as an EPT richness of 8 declines (Figure 21). Assuming a goal of protecting epifaunal substrate conditions representing at least sub-optimal epifaunal substrate conditions (sub-optimal epifaunal substrate scores start at 10 or 50% of the maximum score of 20) as a target and, therefore, fixing the epifaunal substrate score at 10, a TP concentration of 107 $\mu\text{g/L}$ would be required to maintain EPT richness, on average, at 8 taxa. Again, the epifaunal substrate target could be adjusted depending on management expectations for epifaunal substrate quality and rehabilitation potential.

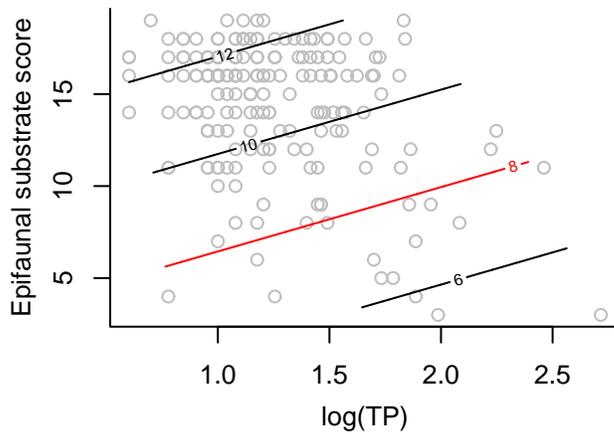


Figure 21. Log(TP) versus epifaunal substrate scores from the least urban MBSS Piedmont ecoregion stream dataset. Contours indicate predicted mean values of EPT richness from the multiple linear regression model. Units are $\mu\text{g/L}$ for log(TP).

3.3 Identifying thresholds from stressor-response relationships

In some cases, characteristics of the estimated stressor-response relationship can inform the selection of candidate criteria. In the relationship shown in Figure 22, a threshold in the relationship occurs when nutrient concentrations reach the point indicated by the red arrow. For example, species richness may remain stable through replacement of intolerant species with tolerant species until concentrations exceed the highest tolerances, beyond which species richness begins to decline. Values of the response variable beyond this point decline steeply. Candidate criteria could be defined above or below this threshold point, depending on the degree of confidence one had in the estimated relationship and the degree to which one wanted to protect against loss of the designated use. In this section, methods for performing such analyses are described.

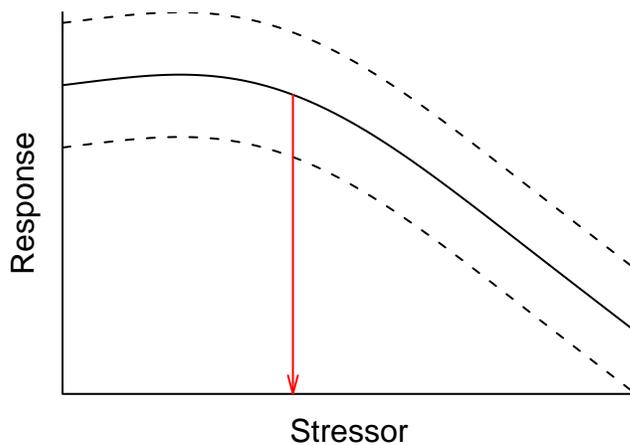


Figure 22. Stressor-response relationship showing threshold behavior. Red vertical arrow indicates a possible criterion value.

3.3.1 Non-parametric changepoint analysis (nCPA)

Description and Data Requirements

Non-parametric changepoint analysis (nCPA) refers to a suite of non-parametric methods used for identifying thresholds or changepoints in bivariate relationships. When a dependent variable is related to an independent variable and scatter plots suggest that a threshold or change in the statistical attributes of the dependent variable exists with increases or decreases in the independent variables, changepoint analysis can be used to identify where that changepoint exists (Breiman et al. 1984, Pielou 1984, Qian et al. 2003). In addition to evidence of a changepoint in a plot, there is often a scientific basis to expect thresholds in nature, where systems do not respond linearly to stress, but rather frequently exhibit non-linear responses (e.g., May 1977, Odum et al. 1979, Connell and Sousa 1983, Scheffer et al. 2001, Brenden et al. 2008, Andersen et al. 2009). nCPA has been used for identifying thresholds associated with nutrient stressors and various responses in freshwaters including effects on plants and invertebrates (King and Richardson 2003, Qian et al. 2003, 2004).

nCPA is used to find some point in the relationship between x and y , where some statistical attribute of y (e.g., the mean or variance) is most different above some point x than below it. In the hypothetical figure below (Figure 23), the mean and variance of Y below the value c are different than above it. The significance of this changepoint can be evaluated and it is presented as either the single changepoint value, a single value with an empirically derived confidence interval, or as a distribution of changepoints based on resampling (King and Richardson 2003, Qian et al. 2003).

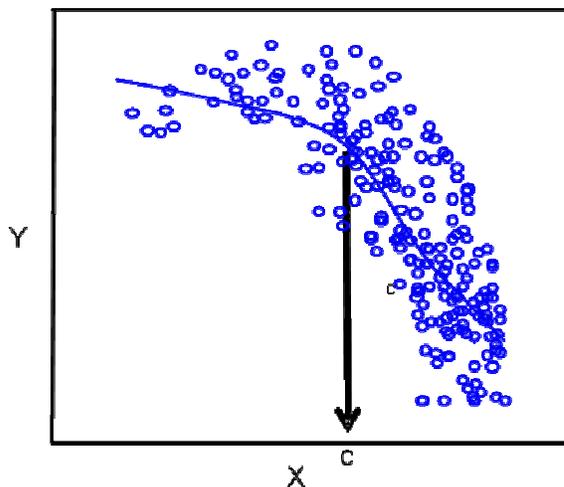


Figure 23. Hypothetical relationship between an independent variable X and a dependent variable Y . The value c indicates a changepoint in the relationship between X and Y , characterized by a change in the mean of Y on either side of c .

Functionally, changepoint analysis is conducted by ordering observations along a stressor gradient (X) and determining which point along that gradient splits the response variable into the two groups with the greatest difference in statistical attribute (e.g., mean, deviance or variance)(King and Richardson 2003). As mentioned earlier, there are different methods for determining changepoints depending on the statistical attribute that is evaluated. Changepoint analysis, as described in this context, will refer to the deviance reduction method based on mean values (King and Richardson 2003, Qian et al. 2003), an abbreviated version of the classification and regression tree (CART) methodology of Breiman et al. (1984). Deviance is the within-group variance of the response variable Y (using means for the group rather than the overall mean). nCPA works by moving iteratively across the entire range of observed stressor values and calculating a deviance estimate for each value of X . The changepoint is defined as that point that maximizes the deviance reduction (Δ_i):

$$\Delta_i = D - (D_{\leq i} + D_{> i}) \text{ (Breiman et al. 1984),}$$

where D = deviance for the data as one group, $D_{\leq i}$ is the deviance of y below and including x_i , and $D_{> i}$ is the deviance of y above x_i along an ordered predictor variable x_1, \dots, x_n , and i is an index for the (Y, X) pairs ordered from smallest to largest on the X variable. The changepoint is the value of x that maximizes Δ_i . This changepoint splits the data into two groups with the greatest difference in the mean value.

Uncertainty in the changepoint location can be quantified with resampling techniques. For example, bootstrap resampling has been used to estimate uncertainty around changepoint estimates, as well as cumulative probability curves (King and Richardson 2003). Also, χ^2 tests have been used to test whether changepoints are statistically significant (King and Richardson 2003), although these authors argued that uncertainty around the changepoint was more important and they only applied the χ^2 test when cumulative distribution patterns for changepoints were wide.

Changepoint analyses are not subject to the same restrictive assumptions of linear regression as deviance reduction is a non-parametric analysis, and can therefore be used to effectively model nutrient and response variables whose underlying relationship is non-linear and/or whose residual variance differs across the range of modeled values (King and Richardson 2003). For linear relationships and relationships characterized by wedge-shaped factor ceilings (Carter and Fend 2005), which are also common with stressor-response relationships in water quality analysis, the linear modeling techniques described above are the more appropriate approaches to use (King and Richardson 2003). Preliminary visual analyses will likely indicate which relationships may yield non-linear changepoints.

Example Applications

In this first example, a changepoint was identified in the relationship between chlorophyll *a* concentrations and TP in the EMAP Northeast Lakes Survey dataset. It appeared from visual inspection that the relationship between these two variables was not linear, but rather chlorophyll increased slowly with TP and then increased dramatically at some TP threshold (Figure 24). nCPA was conducted on these data to evaluate whether a significant changepoint existed. The deviance reduction method and bootstrap resampling were used to estimate the changepoint and empirical confidence intervals around the changepoint, respectively.

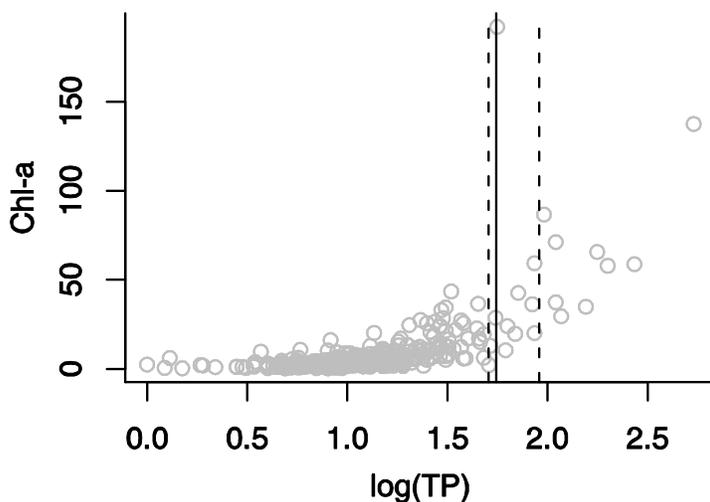


Figure 24. Plot of chlorophyll *a* versus log total phosphorus in EMAP Northeast Lakes Survey sites. The solid vertical line is the changepoint estimate and the dashed lines represent 95 % confidence intervals around the changepoint estimate. Units are $\mu\text{g/L}$ for $\log(\text{TP})$ and chlorophyll *a*.

A changepoint was identified at 55 $\mu\text{g/L}$ TP, above which chlorophyll *a* concentrations increased rapidly with additional phosphorus. The confidence interval around the changepoint was large, 50 – 90, and a χ^2 test gave evidence ($p < 0.05$) of a significant changepoint. Constraining the minimum group size required to identify a group split can be used to reduce the influence of specific values which can reduce the variance around the changepoint (R. King, pers. comm.). Where multiple changepoints exist, regression trees can be used to identify each subsequent splits (King and Richardson 2003).

The second example is from the EMAP-West Xeric region dataset. The relationship between TN and total macroinvertebrate species richness exhibited a non-linear pattern and was evaluated for a potential changepoint using nCPA analysis (Figure 25). In this case, a changepoint was identified at $\log(\text{TN})$ of 3.0 (1000 $\mu\text{g/L}$ TN). The 90th percent confidence interval around this estimate was approximately 250 $\mu\text{g/L}$ to 2.2 mg/L TN.

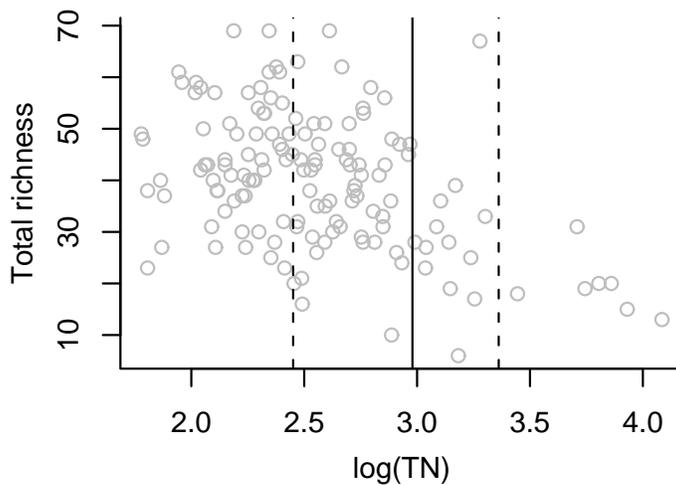


Figure 25. Log(TN) vs. total species richness in EMAP-West Xeric region streams. Solid vertical line indicates changepoint estimate. Dashed vertical lines indicate 90% confidence intervals for changepoint estimated using bootstrap resampling. Units are $\mu\text{g/L}$ for log(TN).

3.3.2 Discontinuous regression models

Description and Data Requirements

Discontinuous regression models, including piecewise or breakpoint regression (also split-point, join point or hockey-stick regression), are a suite of non-linear estimation techniques that can be used to evaluate locations in bivariate plots where the linear relationship between the dependent and independent variable changes. This modeling approach has been used widely in environmental applications (Barrowman and Myers 2000, Ryan and Porth 2007, Brenden and Bence 2008), and has also been applied for nutrient-response analyses (e.g., Dodds et al. 2002).

Discontinuous regression models test whether a specific value of the independent variable might split a single model into two regression models (one above the value and one below the value) that better minimize the loss function (e.g., sum of squared deviations) associated with a set of stressor-response data. In this case, the user may want to test whether a specific potential threshold or breakpoint exists; on either side of which, relationships between the stressor and response differ. In addition to testing specific values, discontinuous regression can also search iteratively through the data to fit a point that minimizes the loss function (Barrowman and Myers 2000). One can also perform this analysis as a non-linear regression where the functional form changes at the breakpoint (e.g., 2 straight lines that are continuous at the breakpoint but have a discontinuous slope). There are a variety of methods. The application for threshold identification in a nutrient criteria context used a quasi-Newton method (Dodds et al. 2002), but other search methods have been used including grid search methods to identify globally best solutions (Barrowman and Myers 2000, Seber and Wild 2003).

Once a point is found that minimizes the loss function on either side of a point, one can evaluate the model fit for each line using standard SLR model fit tests (e.g., model significance, coefficient of determination, and residual plots). One can also test whether the resulting lines are significantly different, which is important for determining whether a piecewise or split-point model is an improvement over a single linear or non-linear model (Kim et al. 2000, Tiwari et al. 2005, UCLA Academic Technology Services 2009). Other variations exist and alternative models, including quantile piecewise regression, can also be used (Brenden et al. 2008).

Example Application

Total species richness and TP data from the entire EMAP-West Stream Survey dataset (as opposed to the Xeric region used in previous examples) were used. The relationship between these variables was fit using an iterative piecewise regression routine to identify a split-point in the relationship between TP and total species richness (Figure 26).

The routine used in this example employed a Gauss-Newton search method to identify the breakpoint that minimizes the standard sum of squared deviations. A breakpoint was identified at a $\log(\text{TP})$ value of 1.0 (10 $\mu\text{g/L}$). Error around breakpoints identified with piecewise regression can be estimated using resampling techniques, as described for other analyses.

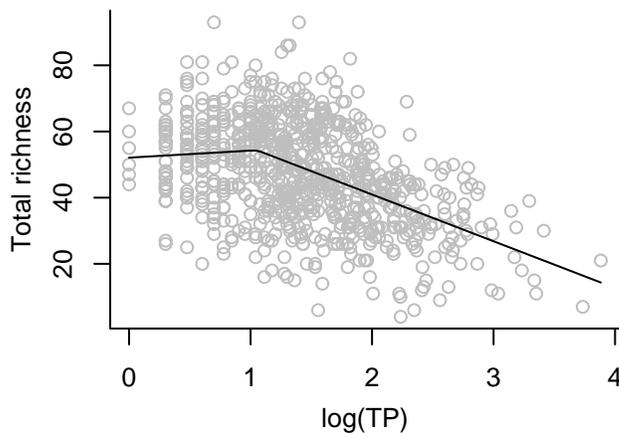


Figure 26. $\log(\text{TP})$ vs. total species richness for EMAP-West Stream Survey data. Line indicates fit from piecewise linear model. Units are $\mu\text{g/L}$ for $\log(\text{TP})$.

4 Evaluating estimated stressor-response relationships

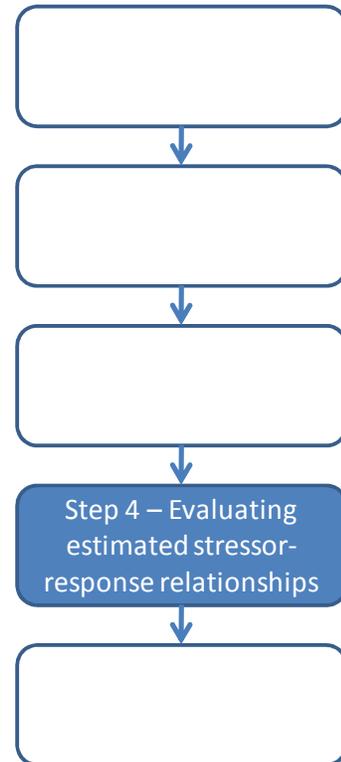
A number of different approaches can be used to develop estimates of stressor-response relationships, and so the next step (Step 4, Figure 1) in deriving numeric nutrient criteria is to evaluate and compare different estimates of the stressor-response

relationship. The predictive performance of different models can be validated (4.1), and the models can be qualitatively assessed in terms of different sources of uncertainty (4.2). The final objective is to select a single stressor-response model for each response variable that best represents the data (4.3).

4.1 Model validation

One approach for validating model predictive performance is to use the estimated model to predict responses in independent validation data (i.e., data not used to calibrate the original relationship). Independent validation data can be acquired in a number of ways, including resampling the study area using the same protocols used when collecting the calibration data. However, this approach may be impractical due to resource constraints. Alternatively, a subset of the original dataset can be held out of the original analysis and reserved for model validation. Different approaches for holding out data have been proposed. One common and robust approach for evaluating the predictive accuracy of the model is to randomly select samples from the study area of interest, and reserve these as validation data. The stressor-response model can then be calibrated with the remaining data and applied to the validation data. When using stressor-response relationships to derive nutrient criteria, one concern is that estimated relationships will be confounded by unsampled or unmodeled factors that are correlated with the chosen nutrient variable. If a validation dataset is specified by randomly selecting samples from the original dataset, it is likely that the covariance structure of the validation data will be similar to that of the calibration data, and the validation test may be misleading. Thus, a complementary approach to randomly reserved validation data is to hold out a *non-random* subset of the original dataset. For example, in a multi-State dataset, one could reserve data from a particular State for subsequent validation studies. Within a given State, one could reserve data from a particular geographic region. This non-random subset will likely have a different covariance structure from the calibration data, and can provide a more rigorous test of the estimated model.

Predictive performance can be quantified by calculating statistics such as a root-mean-square predictive error (RMSPE), which provides an estimate of the average difference between predicted values and observed values.



$$RMSPE = \sqrt{\frac{\sum_{i=1}^N (y_{obs,i} - y_{pred,i})^2}{N}}$$

Where $y_{obs,i}$ is the observed value of the response variable at site i , $y_{pred,i}$ is the model predicted value of the response variable at site i , and N is the number of sites.

Model accuracy will vary substantially depending upon the number of predictive variables included in the model, and the strength with which the response variable is associated with the different predictor variables. However, for the present purposes, it is important to check whether RMSPE calculated for validation data is comparable to that calculated for calibration data. If RMSPE from validation data is comparable to that observed for calibration data, then the model is less likely to be overfit.

To illustrate this approach for model validation, data from Massachusetts was excluded from the EMAP Northeast Lakes Survey dataset, and the SLR model relating $\log(\text{chlorophyll } a)$ and $\log(\text{TP})$ was refit using the remaining data. The model was then used to predict chlorophyll a in Massachusetts. Model predicted values were relatively close to the observed values (1:1 line in Figure 27).

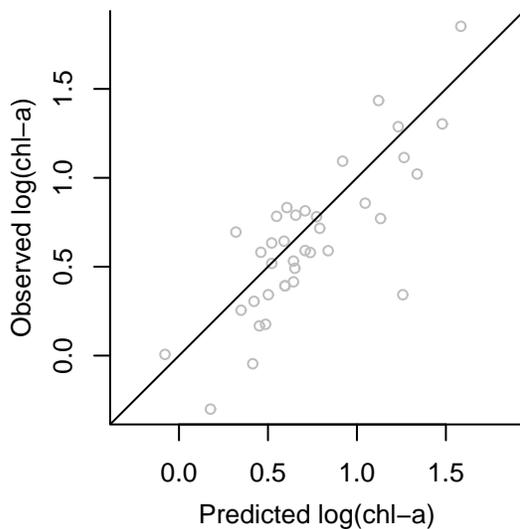


Figure 27. Predicted $\log(\text{chl } a)$ vs. observed $\log(\text{chl } a)$ for EMAP Northeast Lakes Survey sites in Massachusetts. Solid line indicates 1:1 relationship. Units are $\mu\text{g/L}$.

RMSPE calculated for the calibration dataset was 0.28, while RMSPE calculated for the held-out validation data from Massachusetts was 0.27, so the modeled relationship was robust.

4.2 Qualitative assessment of the uncertainty of the estimated stressor-response relationship

Models are by definition an approximate representation of reality, and when using models for decision-making it is prudent to consider model accuracy. Model assumptions should be reviewed, and uncertainty regarding model structure should be considered.

Model assumptions specific to each statistical method have already been considered (Step 3), and the appropriateness of these assumptions should be evaluated with respect to the data. All stressor-response models estimated from cross-sectional or synoptic data must also invoke the assumption that spatial differences in sites can be substituted for temporal differences without a substantial degradation of model accuracy (i.e., the space-for-time substitution). More specifically, stressor-response relationships will often be developed using data collected from many different sites within a State or study region, but when these stressor-response relationships are used to inform nutrient criteria, they are interpreted in terms of their effects on single sites. Space-for-time substitutions have long been used in ecological studies because long-term temporal data are often not available (Fukami and Wardle 2005). In the context of the ecological effects of nutrients, some studies have also shown this assumption to be generally valid (e.g., lake nutrient-chlorophyll *a* models, Jones and Knowlton 2005). If long-term data in a small number of sites are available, analysis of these data can provide support for the validity of the space-for-time substitution.

Uncertainty in model structure should also be considered before finalizing stressor-response models. Questions to consider include whether all appropriate environmental factors have been included in the model, and whether the specified functional form for each variable is appropriate. Unmodeled environmental factors that co-vary with the nutrient variable and affect the response variable on their own can influence the estimated effects of the nutrient variable. Some of these variables can be identified during the initial data exploration phases of the analysis (Step 1). When data for these variables are available, their potential influence on the estimated effects of nutrient concentrations can be quantified by including them as explanatory variables in a multiple linear regression. Propensity score analysis also provides a robust means of incorporating these variables into estimates of stressor-response relationships. The effects of variables for which data are not available can only be considered qualitatively. Conceptual models specifying linkages between human activities, stressor variables, and response variables (Step 2) can provide insights into possibly important, but unsampled variables.

The functional forms assigned to the nutrient variable and to other explanatory variables should also be considered. Simple and multiple linear regressions assume that these function forms are linear (i.e., straight lines), and often this assumption provides a sufficiently realistic representation of the relationship between the nutrient and

response variables. Fitting loess regression and other non-parametric smoothing regressions (Step 1) that relax the assumption of linearity can provide some indication as to whether linear models are appropriate.

4.3 Select the stressor-response model

Six approaches have been described for modeling stressor-response relationships, and one or more of these methods may have been applied to a specific response variable. These methods differ in the assumptions they impose on the residual distribution, the functional form of the stressor response relationship, and their data requirements (Table 1). At this point in the analysis, one should review the assumptions associated with each method and select the model that is most appropriate for the available data.

Of these methods, simple and multiple linear regressions impose the strictest assumptions regarding the data, but when appropriately applied, these models provide the strongest means of inferring the effects of different criteria. Residual variance should be normally distributed (i.e., Gaussian) and constant over the range of predictor variables. This assumption is particularly important when inferring probabilities of impairment from fitted relationships. Relationships are also assumed to be linear. Each of these assumptions can be tested using regression diagnostic plots, or by fitting alternate, less restrictive models.

Table 1. Summary of different method characteristics

| Method | Residual distribution | Functional form | Data requirements |
|----------------------------|--------------------------------|------------------|-------------------|
| Simple linear regression | Constant variance Gaussian | linear | moderate |
| Logistic regression | Binomial | sigmoidal | moderate |
| Quantile regression | none | linear | high |
| Multiple linear regression | Constant variance, Gaussian | linear | high |
| Change point analysis | None | Step change | moderate |
| Discontinuous regression | Constant variance Gaussian | Piecewise linear | high |

Plotting residual values against predicted mean values can provide an indication as to whether residual variances are constant. Alternatively, quantile regression can be used to estimate upper and lower percentiles (e.g., the 95th and 5th percentiles) of the

response variable, as a function of the nutrient variable. Then, the distance between these two quantiles provides an alternate indication of whether residual variances are constant. When residual variance changes substantially, quantile regression can also provide an alternate means of estimating candidate criteria.

The first four methods listed in Table 1 assume that the stressor-response function is a straight line (i.e., linear). With logistic regression, this straight line is a logit-transformed function such that the final curve describing the probability of impairment is sigmoidal. In cases in which a linear model is not appropriate, split point regression can represent relationships that feature a single change in slope. Changepoint analysis is even more flexible as it inherently only assumes that the distribution of mean response values changes in some unspecified manner along the range of predictor values (see Section 3.3). Note that most of the linear methods can be extended to curvilinear functional forms, either by transforming the predictor variables or by fitting non-parametric curves to the data (e.g., Wood and Augustin 2002).

Data requirements for these methods are generally comparable. They require paired observations of stressor and response variables meeting the assumptions described for each analysis. Quantile regression requires more data to estimate extreme quantiles (e.g., 5th and 95th percentiles), and the data requirements for multiple linear regression increase with additional explanatory variables.

In the examples based on EMAP Northeast Lakes Survey data, SLR explained nearly the same amount of variability as MLR, and quantile regression indicated that the residual variance of chlorophyll *a* was reasonably constant across different values of log(TP). Therefore, basing nutrient criteria on SLR seems appropriate.

The example based on macroinvertebrate data from EMAP-West Xeric region stream provided a more challenging choice of stressor-response model. Quantile regression of the bivariate relationships suggests that SLR assumptions were not appropriate. Also, MLR identified a significant influence of other explanatory variables on the effects of TN, and controlling for these other covariates changed the estimates of the effects of TN on macroinvertebrate species richness. Because multiple variables contribute to the observed values of richness, the MLR model likely provides the most accurate estimates of nutrient effects.

In both of these examples, it has been assumed that a threshold value for the response variable is available, and so, one of the first four methods in Table 1 have been selected to translate this response variable threshold to a numeric nutrient criterion. In cases in which response thresholds are not available, use of the last two methods in Table 1 is recommended.

5 Evaluating candidate stressor-response criteria

The process of using a stressor-response relationship to translate an existing response threshold to a comparable nutrient criterion inherently requires that the values of response variables are *predicted* at lower nutrient concentrations. That is, a criterion should be set such that when nutrient concentrations are reduced to this value at sites at which the criterion is exceeded, the values of the response variable at these sites meet management objectives (e.g., a certain proportion of sites maintain designated uses). In this last step of the process, step 5 (Figure 1), an approach is outlined for explicitly predicting conditions that might result after implementing nutrient criteria. These predictions can then be used to evaluate candidate criteria.

As described in step 3, for EMAP Northeast Lakes Survey data, lake chlorophyll *a* concentrations are significantly associated with TP concentrations (Figure 28). Following the methods outlined in step 3, a criterion to maintain chlorophyll *a* at 15 µg/L in approximately 50% of lakes could be set by calculating the point at which the chlorophyll *a* concentration predicted by the mean simple linear regression line is the same as the desired biological threshold. The value of this candidate criterion is $\log(\text{TP}) = 1.6$ (TP = 44 µg/L) (red arrow in Figure 28).

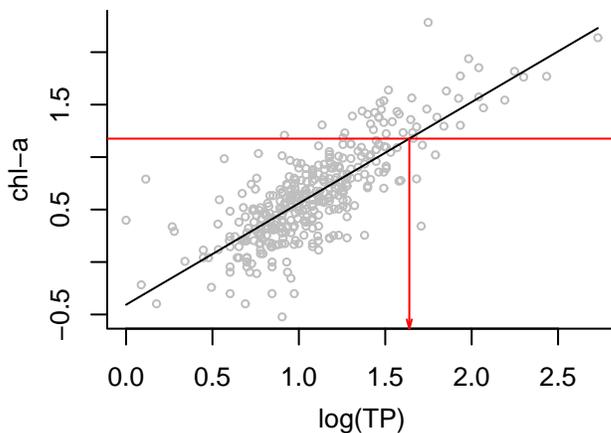
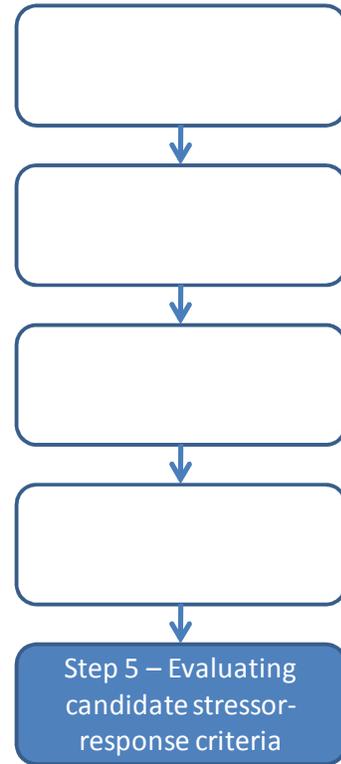


Figure 28. Simple linear regression estimate of stressor-response relationship. Red horizontal line indicates a possible threshold value for chlorophyll *a* of 15 µg/L. The red arrow indicates a candidate nutrient criterion. Units are µg/L.

Inherent in this use of the stressor-response model is a prediction of the potential change in the value of the response variable at sites at which $\log(\text{TP})$ exceeds the

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

candidate criterion. More specifically, imagine that lakes in which the candidate criterion was exceeded were managed such that TP was reduced to the criterion value. The new chlorophyll *a* in these managed lakes could be predicted by projecting chlorophyll *a* values from the current nutrient concentration back to the criterion value, using the estimated stressor-response relationship. An example of this prediction for one lake is shown as a blue arrow in Figure 29, where the slope of the arrow is identical to the slope estimated from the simple linear regression, and the arrow extends from the observed values of log(TP) and chlorophyll *a* to the candidate criterion value for log(TP) and a predicted value of chlorophyll *a*.

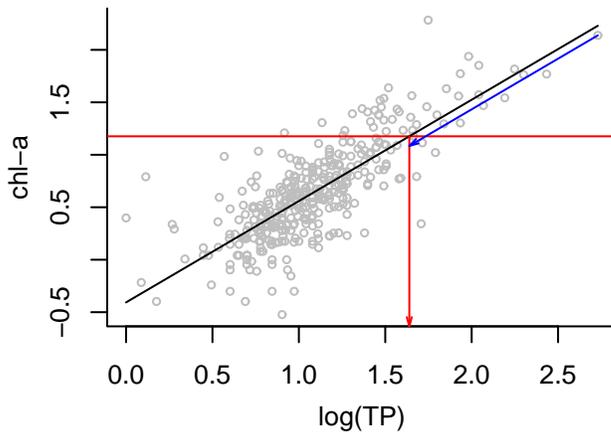


Figure 29. Example of using an estimated stressor-response relationship to project the effects of criterion. Units are $\mu\text{g/L}$ for log(TP) and log(chl-a).

This process can be repeated for all sites that exceed each candidate criterion, as illustrated in Figure 30. (A candidate criterion value of log TP = 2 is assumed here to reduce the number of arrows on the figure and improve plotting clarity). Then, after predicting chlorophyll *a* concentrations for these sites, the new distribution of chlorophyll *a* values (Figure 31) across the dataset can be examined with respect to management targets for the biological response.

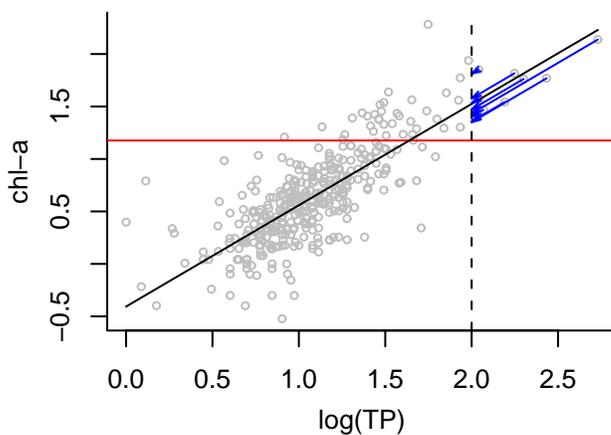


Figure 30. Example of projecting chlorophyll *a* in lakes to a candidate criterion value of log(TP) = 2 (TP = 100 $\mu\text{g/L}$). Units are $\mu\text{g/L}$ for log(chl-a).

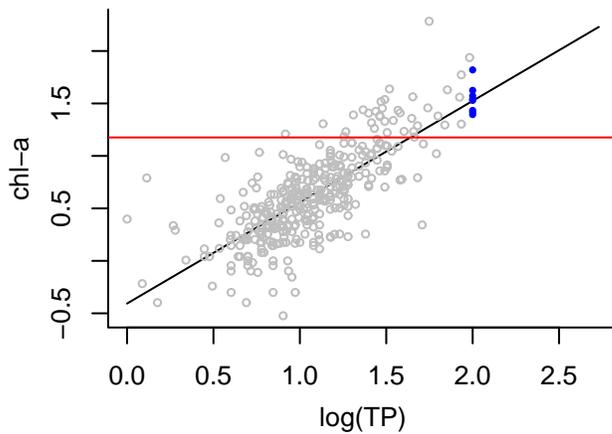


Figure 31. Resulting distribution of chlorophyll a after projection. Solid blue symbols show projected chlorophyll a concentrations at sites where original $\log(\text{TP})$ was greater than 2. Units are $\mu\text{g}/\text{L}$ for $\log(\text{TP})$ and $\log(\text{chl-a})$.

Where the assumptions associated with predicting conditions at lower nutrient concentrations are consistent with assumptions inherent to using stressor-response relationships for setting criteria, these predictions can be used to assess the potential effects of candidate criteria. One way to use these predictions to evaluate the effects of candidate nutrient criteria is to calculate the predicted probability of exceeding the biological threshold in sites that originally exceeded the candidate nutrient criterion. In Figure 30, eight lakes exceeded the candidate criterion of $\log(\text{TP}) = 2$. After the candidate criterion is implemented, projected chlorophyll a concentrations in these 8 lakes still exceeded the targeted biological threshold (Figure 31). Thus, the predicted probability of impairment is still 100% even after implementation of this candidate nutrient criterion. The predicted probability of impairment defined here is exactly analogous to the conditional probability of impairment described in step 1, and the two values can be shown on the same plot to compare *observed* conditions with *predicted* conditions after the implementation of different candidate criteria (Figure 32).

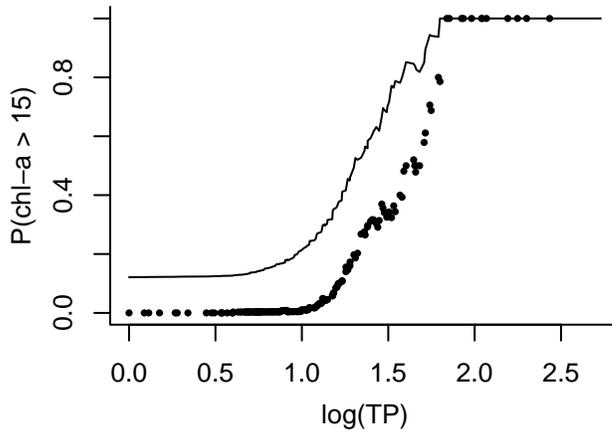


Figure 32. Observed and predicted probabilities of impairment for different nutrient concentrations. Solid line: observed probability of impairment. Solid symbols: predicted probability of impairment. Units are $\mu\text{g/L}$ for $\log(\text{TP})$.

The predicted probabilities of impairment are 0% for nutrient criteria less than approximately $\log(\text{TP}) = 1$ ($\text{TP} = 10 \mu\text{g/L}$), while predicted probabilities of impairment are 100% for criterion values greater than about 1.8 ($\text{TP} = 63 \mu\text{g/L}$) (Figure 32). In between these two limits, different predicted probabilities of impairment are associated with different candidate criteria. Examining the predicted effects of candidate criterion in this way provides information that is more directly applicable to risk management decisions. That is, for any criterion value one can state the probability of impairment for sites at which current nutrient concentrations exceed the candidate criterion value.

One further advantage of using this approach to evaluate candidate criteria is that the framework expands easily to incorporate more involved analyses of the stressor-response relationship. So, this approach can be applied regardless of the type of model selected for estimating stressor-response relationships. Stressor-response relationships for the EMAP West Xeric region streams provide a case in which the effects of increased nutrients are best estimated using MLR (Figure 33). In this case variability about the mean relationship between total richness and TN is large, and reducing TN concentrations does not appear to strongly affect the number of sites that attain the specified response threshold.

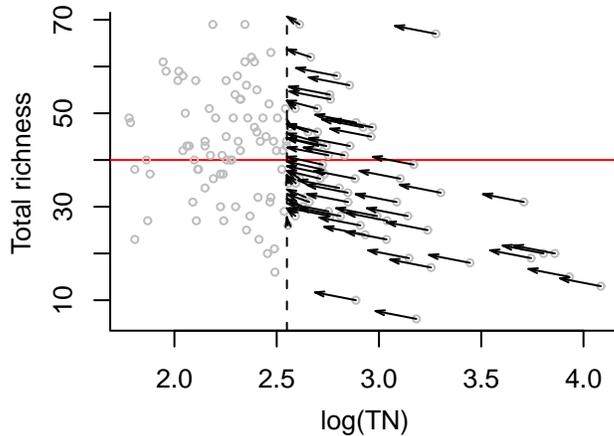


Figure 33. Total species richness vs. log(TN) for EMAP-West Xeric region streams. Arrows indicate predicted direction of change from multiple linear regression. Red horizontal line indicates possible response threshold, and dashed vertical line indicates possible criterion value. Units are $\mu\text{g/L}$ for log(TN).

Plots of predicted probabilities of impairment are consistent with qualitative observations from the scatter plot (Figure 34). Candidate criteria that are greater than approximately $\log(\text{TN}) = 2.7$ ($\text{TN} = 500 \mu\text{g/L}$) do not substantially change the probability of observing impaired sites, while candidate criteria between $\log(\text{TN}) = 2$ and 2.7 ($\text{TN} = 100 - 500 \mu\text{g/L}$) reduce the probability of impairment by about 10%. Candidate criteria less than 2 have more marked effects on the probability of impairment. In this case, stressor-response analyses indicated that MLR provided the most accurate estimates of the effects of TN, but different criterion values were identified by each method. The predictive summary shown in Figure 34 provides a means of evaluating each candidate criterion value with respect to predicted probabilities of impairment.

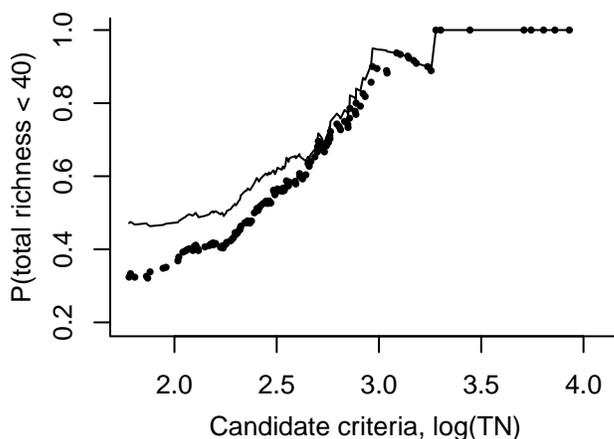


Figure 34. Observed and predicted probabilities of impairment for EMAP-West Xeric region streams. Predicted probabilities of impairment based on a MLR model (solid black symbols). Units are $\mu\text{g/L}$ for log(TN).

In addition to the assumptions discussed above, several different aspects of model uncertainty can influence the predictions calculated from stressor-response relationships. First, the slope of the estimated stressor-response relationship is uncertain. One estimate of this uncertainty is the standard error for the regression coefficient provided. This standard error could be incorporated into the predictions for each point, but calculation of this standard error assumes that the model is correctly specified. Unmodeled factors that are correlated with the nutrient variable and influence the response variable on their own can add further uncertainty to the estimated regression coefficient that would not be represented by the estimated standard error. Thus, a qualitative assessment of the accuracy and potential biases of the estimated model (as discussed in step 4) is critical. For example, if a qualitative assessment of the value of the regression coefficient suggests that the effects of nutrients are over-estimated (i.e., the magnitude of the regression coefficient is too large), then projected values of total richness would be higher than would be observed. This uncertainty may suggest that a lower criterion value would be appropriate.

Predictions based on the estimated slope of the stressor-response relationships are also subject to assumptions regarding the validity of space-for-time substitution and assumptions regarding the nature of management actions. As noted in step 4, stressor-response relationships are likely to be estimated using data collected across many different sites. Then, in setting candidate criteria, one assumes that this relationship is applicable to individual sites, where nutrient concentrations change over time. In predicting only the effects of reductions in nutrients, one also assumes that management actions only influence nutrient concentration. In reality, most management actions would likely affect other stressors as well. For example, increasing the extent of riparian buffers is known to reduce both TP loading and inputs of fine sediment. Thus, projecting conditions based only on changes in nutrient concentration may underestimate the effects of certain management actions. Finally, estimates of the effects of nutrients are based on data in which other factors co-vary with nutrient concentrations. When nutrient concentrations are reduced to calculate predictions, the combination of the decreased nutrient concentrations with other environmental factors may diverge from combinations of nutrients and other factors that are observed in the dataset. Thus, confirming that predicted conditions are still within the range of observed conditions can help protect against extrapolating beyond the scope of the data.

The effects of both unmodeled factors that are not correlated with the nutrient variable and sampling variability of the response variable are also important. In Figure 33, observed values of total richness are distributed widely about the mean regression line because of these two sources of variability. This residual variability influences the percentage of sites that exceed the total richness threshold and ultimately, influences the perceived effectiveness of the candidate criterion in helping to reach management objectives. Ideally, residual variability would be explicitly modeled when predicting conditions at lower nutrient concentrations. However, the simpler approach adopted

Empirical Approaches for Nutrient Criteria Derivation SAB Review Draft

here takes this variability into account by projecting the value for each sample from the original, observed value for that sample. That is, the existing residual variability in the observations is maintained in the predictions at lower nutrient concentrations. This approach assumes the magnitude of residual variability is constant across the range of modeled conditions, an assumption that has already been invoked when fitting the original regression model. Note that because of residual variability, a model prediction for a single site should not be interpreted as an accurate projection of future conditions at that particular site. The uncertainty for predictions for individual sites is at least as large as the observed residual variability, so it is very difficult to accurately predict future conditions at a single site. However, when errors are normally distributed the set of predictions for all sites can provide a reasonably accurate estimate of the distribution of conditions after the implementation of a candidate criterion.

The list of assumptions and uncertainties described here is extensive, but as noted earlier, these issues are present regardless of whether predictions are explicitly calculated and examined. That is, when stressor-response relationships are used to set candidate criteria, predictions of conditions at lower nutrient concentrations are implicitly assumed. Therefore, evaluating criteria in terms of predicting conditions provides a means of more clearly understanding and communicating sources of uncertainty in the analysis.

Criteria based on the characteristics of the stressor-response relationship (e.g., change points) may be more challenging to evaluate. If regression models that represent the stressor-response relationship are available from other analyses, and a response threshold has been defined, then the approach outlined here can be applied. However, in many cases, criteria may be based on characteristics of the stressor-response relationship because response thresholds are not available and/or regression models cannot be fit. In these cases, evaluation of the appropriateness of candidate criteria is more subjective. A threshold behavior (e.g., Figure 26) may indicate that some assimilative capacity for excess nutrients is available in the system, and therefore, setting the criterion at this threshold may be particularly appropriate. Evidence of the same type of response in other studies may provide support for the selected criteria.

References

Allan, J.D. 1995. *Stream Ecology: Structure and Function of Running Waters*. Chapman & Hall, New York.

Andersen, T., J. Carstensen, E. Hernandez-Garcia, and C.M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology and Evolution* (in press).

Ashby, D. 2006. Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine* 25: 3589-3631.

Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15:185-211.

Barrowman, N.J. and R.A. Myers. 2000. Still more spawner-recruitment curves: the hockey stick and its generalizations. *Canadian Journal of Fisheries and Aquatic Sciences* 57: 665-676.

Barry, D. and J.A. Hartigan. 1993. A bayesian analysis for change point problems. *Journal of the American Statistical Association* 88: 309-319.

Breiman, L., J.H. Friedman, R. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.

Brenden, T.O. and J.R. Bence. 2008. Comment: use of piecewise regression models to estimates changing relationships in fisheries. *North American Journal of Fisheries Management* 28: 844-846

Brenden, T.O., L. Wang, and Z. Su. 2008. Quantitative identification of disturbance thresholds in support of aquatic resource management. *Environmental Management* 42:821-832.

Bryce, S.A., G.A. Lomnický, P.R. Kaufmann, L.S. McAllister, and T.L. Ernst. 2008. Development of biologically based sediment criteria in mountain streams of the western United States. *North American Journal of Fisheries Management* 28: 1714-1724

Cade, B.S. and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1:412-420.

Carlson R.E. 1977. A trophic state index for lakes. *Limnology and Oceanography* 22:361-369.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Carter, J.L. and S.V. Fend, 2005. Setting Limits: The Development and Use of Factor-Ceiling Distributions for an Urban Assessment Using Macroinvertebrates. In: Effects of Urbanization on Stream Ecosystems. L.R. Brown, R.H. Gray, R.M. Hughes, and M.R. Meador (editors). American Fisheries Society Symposium 47, Bethesda, Maryland, pp. 179-192.

Chapra, S.C. 1997. Surface Water-Quality Modeling, McGraw-Hill, New York, N.Y.

Chevan, A., and M. Sutherland. 1991. Hierarchical partitioning. The American Statistician 45:90-96.

Cleveland, W.S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74:829-836

Cleveland, W.S. 1993. Visualizing Data. Summit, New Jersey, Hobart Press.

Cleveland, W.S. and S.J. Devlin. 1988. Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American Statistical Association 83:596-610

Cleveland, W.S., E. Grosse and W.M. Shyu. 1992. Local Regression Models. Statistical Models in S.J.H. Chambers and T. J. Hastie. Pacific Grove, CA, Wadsworth & Brook: 309-376.

Cochran, W G. 1977. Sampling Techniques. New York, NY: John Wiley & Sons.

Connell, J.H. and W.P. Sousa. 1983. On the evidence needed to judge ecological stability or persistence. American Naturalist 121: 789–824.

Cross, W.F., J.B. Wallace, A.D. Rosemond and S.L. Eggert. 2006. Whole-system nutrient enrichment increases secondary production in a detritus-based ecosystem. Ecology 87:1556-1565.

Dodds, W.K., V.H. Smith, and K. Lohman. 2002. Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. Canadian Journal of Fisheries and Aquatic Sciences 59: 865-874.

Dodds, W.K., V.H. Smith and K. Lohman. 2006. Erratum: Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. Canadian Journal of Fisheries and Aquatic Sciences 63:1190-1191.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Dodds, W.K., V.H. Smith, and B. Zander. 1997. Developing nutrient targets to control benthic chlorophyll levels in streams: A case study of the Clark Fork River. *Water Research* 31:1738-1750.

Elwood, J.W., J.D. Newbold, A.F. Trimble, and R.W. Stark. 1981. The limiting role of phosphorus in a woodland stream ecosystem: effects of P enrichment on leaf decomposition and primary producers. *Ecology* 62:146–158.

Feminella, J.W. and C.P. Hawkins. 1995. Interactions between stream herbivores and periphyton: A quantitative analysis of past experiments. *Journal of the North American Benthological Society* 14: 465–509.

Florida Department of Environmental Protection (FDEP). 2008. Technical support document: Derivation of a Revised Specific Conductance Criterion for Florida Class I & III Freshwaters. December 2008 draft. Florida Department of Environmental Protection, Bureau of Assessment and Restoration Support, Tallahassee, FL.

Fukami, T. & Wardle, D.A. 2005. Long-term ecological dynamics: reciprocal insights from natural and anthropogenic gradients. *Proceedings of the Royal Society B: Biological Sciences* 272: 2105-2115.

Greenland, P. and R.O. Bonow. 2008. How Low-Risk is a Coronary Calcium Score of Zero? The Importance of Conditional Probability. Editorial. *Circulation* 117: 1627-1629.

Greenland, S. 2000. Causal analysis in health sciences. *Journal of the American Statistical Association* 95: 286 – 289.

Gröger, J.P., H. Winkler, and R.A. Rountree. 2007. Population dynamics of pikeperch (*Sander lucioperca*) and its linkage to fishery driven and climatic influences in a southern Baltic lagoon of the Darss-Zingst Bodden Chain. *Fisheries Research* 84: 189-201.

Halliwell, D.B., T.R. Whittier and N.H. Ringler. 2001. Distributions of lake fishes of the northeast USA - III. Salmonidae and associated coldwater species. *Northeastern Naturalist* 8(2): 189-206.

Harrell Jr, F.E., K.L. Lee, and D.B. Mark. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 28:361-387.

Heino, J., T. Muotka, and R. Paavola. 2003. Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. *Journal of Animal Ecology* 72: 425–434.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Heiskary, S. and W.W. Walker. 1988. Developing phosphorus criteria for Minnesota lakes. *Lake and Reservoir Management* 4(1):1-9.

Henry, S.G.C. 1969. Alcohol, drinking patterns, and the psychological probability of success. *Behavioral Science* 14: 19-27.

Hill, A.B. 1965. The environment and disease: Association or causation. *Proceedings of the Royal Society Medicine – London*. 58:295 – 300.

Hillebrand, H. 2002. Top-down versus bottom-up control of autotrophic biomass- A meta-analysis of experiments with periphyton. *Journal of the North American Benthological Society* 21: 349–369.

Hogg, R.V. and J. Ledolter. 1992. *Applied Statistics for Engineers and Physical Scientists*. Macmillan Publishing Co., New York, New York .

Hollister, J.W., H.A. Walker, and J.F. Paul. 2008. CProb: a computational tool for conducting conditional probability analysis. *Journal of Environmental Quality* 37: 2392-2396.

Imai, K. and D.A. Van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* 99: 854–866.

Jones, J.R. and M.F. Knowlton. 2005. Chlorophyll response to nutrients and non-algal seston in Missouri reservoirs and oxbow lakes. *Lake and Reservoir Management* 21:361 – 371.

Kalff, J. 2001. *Limnology*. Prentice Hall, New York.

Kerans, B.L. and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* 4:768-785.

Kim, H.-Y., M.P. Fay, E.J. Feuer, and D.N. Mithune. 2000. Permutation methods for jointpoint regression with applications to cancer rates. *Statistics in Medicine* 19:335-351.

King, R.S., M.E. Baker, D.F. Whigham, D.E. Weller, T.E. Jordan, P.F. Kazayak, and M.K. Hurd. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications* 15:137-153.

King, R.S. and C.J. Richardson. 2003. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management* 31: 795-809.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Klemm, D.J., K.A. Blocksom, F.A. Fulk, A.T. Herlihy, R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, W.T. Thoeny, M.B. Griffith and W.S. Davis. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highland streams. *Environmental Management* 31:656-669.

Koenker, R. 2005. *Quantile regression*. Cambridge University Press, Cambridge, UK.

Koenker, R. and G. Bassett, Jr. 1978. Regression quantiles. *Econometrica* 46:33-50.

Koenker, R. and K.F. Hallock, 2001. Quantile regression. *Journal of Economic Perspectives* 15:43-156.

Lamon III, E.C. and S.S. Qian. 2008. Regional scale stressor-response models in aquatic ecosystems. *Journal of the American Water Resources Association* 44: 771-781.

May, R.M. 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* 269: 471-77.

MacNally R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – ‘predictive’ and ‘explanatory’ models. *Biodiversity and Conservation* 9: 655-671.

MacNally R. 2002. Multiple regression and inference in ecology and conservation biology: further comments on retention of independent variables. *Biodiversity and Conservation* 11: 1397-1401.

Manly, B.F.J., 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall, New York, New York .

Ming Tan, X.X. 1996. Continuous and group sequential conditional probability ratio tests for phase II clinical trials. *Statistics in Medicine* 15: 2037-2051.

Mitchell, H.G. 1991. A bibliography and comments on the use of statistical models in epidemiology in the 1980s. *Statistics in Medicine* 10: 1819-1885.

Morrice, J.A., N. P. Danz, R.R. Regal, J.R. Kelly, G.J. Niemi, E.D. Reavie, T.P. Hollenhorst, R.P. Axler, A.S. Trebitz, A.M. Cotter, and G.S. Peterson. 2008. Human influences on water quality in Great Lakes coastal wetlands. *Environmental Management* 41:347-357

Odum, E.P., J.T. Finn and E.H. Franz. 1979. Perturbation theory and the subsidy-stress gradient. *Bioscience* 29:344-352.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Paul, J.F. and M.E. McDonald. 2005. Development of empirical, geographically specific water quality criteria: A conditional probability analysis approach. *Journal of the American Water Resources Association* 41: 1211-1223.

Peterson, B.J., L. Deegan, J. Helfrich, J.E. Hobbie, M. Hullar, B. Moller, T.E. Ford, A. Hershey, A. Hiltner, G. Kipphut, M.A. Lock, D.M. Fiebig, V. McKinely, M.C. Miller, J.R. Vestal, R. Ventullo and G. Volk. 1993. Biological responses of a tundra river to fertilization. *Ecology* 74(3):653-672.

Pielou, E.C. 1984. *The interpretation of ecological data: A primer on classification and ordination*. John Wiley and Sons, New York.

Qian, S.S., King, R.S., and C.J. Richardson, C.J. 2003. Two statistical methods for the detection of environmental thresholds. *Ecological Modeling* 166: 87–97.

Qian, S.S., Pan, Y., and R.S. King. 2004. Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. *Ecological Indicators* 4:29-37.

Reckhow, K.H., G.B. Arhonditsis, M.A. Kenney, L. Hauser, J. Tribo, C. Wu, L.J. Steinberg, C. A. Stow, S. J. McBride. 2005. *A Predictive Approach to Nutrient Criteria*. *Environmental Science and Technology*. 39(9): 2913-2919.

Riseng, C.M., M.J. Wiley, and R.J. Stevenson. 2004. Hydrologic disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. *Journal of the North American Benthological Society* 23:309-326.

Rosemond, A.D., P.J. Mulholland, and J.W. Elwood. 1993. Top-down and bottom-up control of stream periphyton: effects of nutrients and herbivores. *Ecology* 74:1264–1280.

Rosenbaum, P.R. 2002. *Observational Studies*, 2nd Edition. Springer, New York.

Rosenbaum, P.R. and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688-701.

Ryan, S.E. and L.S. Porth. 2007. A tutorial on the piecewise regression approach applied to bedload transport data. United States Department of Agriculture, Forest Service, Rocky Mountain Research Station. Fort Collins, CO. General Technical Report RMRS-GTR-189.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

Scheffer, M., D. Straile, E.H. van Nes, and H. Hosper. 2001. Climatic warming causes regime shifts in lake food webs. *Limnology and Oceanography* 46: 1780–83.

Schindler, D.W. 1974. Eutrophication and recovery in experimental lakes: Implications for lake management. *Science* 184: 897-899.

Schindler, D.W., Hesslein, R.H. and Turner, M.A. 1987. Exchange of nutrients between sediments and water after 15 years of experimental eutrophication. *Canadian Journal of Fisheries and Aquatic Sciences* 44 (Suppl. 1): 26-33.

Seber, G.A.F., and C.J. Wild. 2003. *Nonlinear regression*. Wiley, Hoboken, New Jersey.

Shipley, B. 2000. *Cause and correlation in Biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press, Cambridge, UK.

Soranno, P.A., K. Spence-Cheruvellil, R.J. Stevenson, S.L. Rollins, S.W. Holden, S. Heaton, and E. Torng. 2008. A Framework for Developing Ecosystem-Specific Nutrient Criteria: Integrating Biological Thresholds with Predictive Modeling. *Limnology and Oceanography* 53:773-787.

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman and W.A. Brungs. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. PB85-227049. National Technical Information Service, Springfield, VA.

Stoddard, J.L., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, D.V. Peck, S.G. Paulsen, A.R. Olsen, D.P. Larsen, J. Van Sickle, and T.R. Whittier. 2006. Mid-Atlantic Integrated Assessment (MAIA): State of the flowing waters report. EPA-620-R-06-001. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. February 2006.

Stoddard, J.L., D.P. Larsen, C.P. Hawkins, R.K. Johnson and R.H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16(4):1267-1276.

Tiwari, R.C., K.A. Cronin, W. Davis, E.J. Feuer, B. Yu, and S. Chib. 2005. Bayesian model selection for joinpoint regression with application to age-adjusted cancer rates. *Applied Statistics* 54:9129-939.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Reading, Massachusetts, Addison-Wesley Publishing Co.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

UCLA Academic Technology Services . 2009. Stata Topics: Regression. UCLA: Academic Technology Services, Statistical Consulting Group. From <http://www.ats.ucla.edu/stat/stata/topics/regression.htm> (accessed March 24, 2007).

U.S. Environmental Protection Agency (USEPA). 2000a. Nutrient Criteria Technical Guidance Manual: Rivers and Streams. EPA-822-B-00-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

USEPA. 2000b. Nutrient Criteria Technical Guidance Manual. Lakes and Reservoirs. EPA-822-B-00-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2001. Nutrient Criteria Technical Guidance Manual. Estuarine and Coastal Marine Waters. EPA-822-B-01-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2002. Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2006a. Framework for developing suspended and bedded sediments (SABS) water quality criteria. EPA-822-R-06-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2006b. The Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams. EPA 841-B-06-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2008. Nutrient Criteria Technical Guidance Manual. Wetlands. EPA-822-B-08-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Van Dam, H., Mertens, A. & Sinkeldam, J. 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology*, 28 (1): 117-133.

Van Nieuwenhuysse, E.E. and J.R. Jones. 1996. Phosphorus-chlorophyll relationship in temperate streams and its variation with stream catchment area. *Canadian Journal of Fisheries and Aquatic Sciences* 53:99-105

Vollenweider, R.A. 1976. Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication. *Memorie dell'Istituto Italiano di Idrobiologia* 33:53-83.

Vollenweider R.A., Kerekes J.J. 1980. Background and summary results of the OECD cooperative program on eutrophication. In: *Proceedings of an International Symposium*

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

on Inland Waters and Lake Restoration, pp. 26-36. U.S. Environmental Protection Agency. EPA 440/5-81-010.

Wang, L., D.M. Robertson, and P.J. Garrison. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams. *Environmental Management* 39:194–212.

Wetzel, R.G. 2001. *Limnology—Lake and River Ecosystems*, 3rd Edition. Academic Press. New York, N.Y.

Whittier, T.R., D.B. Halliwell and R.A. Daniels. 2001. Distributions of lake fishes in the Northeast - IV: Benthic and small water-column species. *Northeastern Naturalist* 8(4): 455-482.

Whittier, T.R., D.B. Halliwell and S.G. Paulsen. 1997. Cyprinid distributions in Northeast USA lakes: evidence of regional-scale minnow biodiversity losses. *Canadian Journal of Fisheries and Aquatic Sciences* 54(7): 1593-1607.

Whittier, T.R. and T.M. Kincaid. 1999. Introduced fish in northeastern USA lakes: Regional extent, dominance, and effect on native species richness. *Transactions of the American Fisheries Society* 128(5): 769-783.

Wilk, M.B. and R. Gnanadesikan. 1968. Probability Plotting Methods for the Analysis of Data. *Biometrika* 55: 1-17.

Wood, S.N. and N.H. Augustin. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157: 157– 77.

Appendix A – Data Descriptions

EMAP-West Streams Survey: This was the western States component of the USEPA ORD program to conduct a statistically-valid survey of the biological condition of streams throughout the U.S. This project sampled wadeable stream reaches across 12 western States during the summers of 2000 to 2002. The stream reaches were selected for sampling with a probability design. Extensive biological, chemical, physical habitat, and landscape-scale parameters were collected at each sample site. A total of 827 sites were sampled and have the full suite of measurements available. Refer to USEPA (2006b) for details of the survey. More detailed information is also available on the EMAP website (<http://www.epa.gov/emap/>).

EMAP Northeast Lakes Survey: This was an early pilot program of USEPA ORD EMAP project. Data were acquired in 1991-1994 and are available from the EMAP web site. Approximately 330 lakes were sampled over four years. The lakes were selected for sampling with a probability design, and a suite of indicators were collected. Specifics of the survey can be found in Halliwell et al. (2001), Whittier and Kincaid (1999), and Whittier et al. (1997, 2001).

EMAP Mid-Atlantic Integrated Assessment: This was a Mid-Atlantic sampling program of the USEPA ORD EMAP focused on characterizing water quality conditions in streams across the Mid-Atlantic Highlands region (an area that includes the Central Appalachians, the Central Appalachian Ridges and Valleys, and the Blue Ridge Mountains ecoregions). Approximately 850 sites were sampled between 1993 and 1998 and a suite of indicators were collected consistent with most EMAP programs. Specifics of the survey and results can be found in Stoddard et al. (2006).

USGS National Water Quality Assessment (NAWQA) Program: NAWQA is a national water quality monitoring program initiated by the USGS in 1991 as a long-term water resources monitoring network in support of State and federal management needs. NAWQA focuses its efforts on large river basins or study units and sampled 51 study units between 1991 and 2001. NAWQA collects data on macroinvertebrates, fish, algae, water quality, and habitat. They also characterize land use within each watershed draining to their sampling sites. The Susquehanna River Basin Study Unit includes sites in several ecoregions, including the Allegheny Plateau and Ridge and Valley. NAWQA sampling protocols are all available along with detailed reports on the Susquehanna River Basin Study unit at the NAWQA website (<http://water.usgs.gov/nawqa/>).

Maryland Biological Stream Survey (MBSS) Piedmont dataset: The Maryland Biological Stream Survey was started by the Maryland Department of Natural Resources in 1993 and has been sampling streams around the State using a combined randomized and fixed sentinel site network design for the last 16 years. The MBSS samples macroinvertebrates, fish, habitat, and water chemistry using standardized protocols.

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

They also characterize land cover/land use characteristics for the watersheds draining to each site. The program has sampled more than 2000 sites since inception. The Piedmont dataset consists of those sites sampled by MBSS on the Piedmont ecoregion. More detailed information is also available on the MBSS website (<http://www.dnr.md.gov/streams/mbss/>).

Appendix B – Propensity Score Analysis

Propensity score analysis was developed specifically to more accurately estimate stressor-response relationships from observational data (Rosenbaum 2002). This approach provides a means for optimally controlling for the effects of all observed covariates when estimating the effects of nutrients. That is, to best estimate the causal effect of the selected nutrient variable, one would like to identify groups of samples that are similar with regard to covariate distributions, but differ with regard to the nutrient variable. If only a single factor (e.g., substrate sediment) co-varied with the nutrient variable, one could stratify the dataset by this one factor, splitting the dataset into groups with similar values. However, this approach rapidly becomes impractical as the number of factors increases. Propensity scores (Rosenbaum and Rubin 1983, Rosenbaum 2002, Imai and Van Dyk 2004) summarize the contributions of all known covariates as a single variable. A propensity score is estimated by modeling the value of the nutrient variable as a function of covariate values using regression analysis. Then, the predicted mean nutrient concentration in each stream is the propensity score. Stratifying by the propensity score effectively splits the dataset into groups with similar covariate distributions. Once the dataset is stratified, causal effects of nutrients can be more confidently estimated within each group because distributions of other covariates are similar.

To most effectively apply the propensity score approach, extensive data quantifying covariate values, in addition to the nutrient and response variable, must be collected at all sampling locations. These relatively intense data requirements derive from the fact that the analysis cannot control for covariates for which data are not available. Operationally, any environmental factor that is thought to co-vary with nutrient concentrations should be included in the analysis. Examples include measurements of other direct stressors on the aquatic community (e.g., substrate sand/fines, water chemistry), natural factors that determine stream type (e.g., watershed size, geographic location, canopy cover), and land use summaries that provide information regarding the intensity of nearby human activities (e.g., percent agriculture in the catchment). In addition to the large number of different variables, propensity score analysis requires data from a large number of distinct sites because a broad variety of different sites are required to effectively identify groups of sites with similar covariate distributions.

Example Application

USEPA Office of Research and Development (ORD) EMAP West Stream Survey data (Appendix A) are ideally suited for applying propensity score analyses because a broad suite of variables was collected at each site, and many sites were sampled. A generalized additive model (Wood and Augustin 2002) was used to model observed TN concentrations as a function of fourteen covariates (elevation, grazing intensity index, longitude, log annual precipitation, log catchment area, log Cl^- , log HCO_3^- , log SO_4^{2-} ,

percent substrate sand/fines, stream temperature, percent catchment agriculture land use, percent catchment urban land, percent open canopy, and riparian agricultural disturbance). Then, model-predicted mean values of log TN at each site were the propensity scores for each site (Imai and Van Dyk 2004), and the entire dataset was divided into 6 groups based on equally-spaced percentiles of the propensity score.

Table 2. Correlation coefficients between covariates and TN. Correlation coefficients tabulated for full dataset (r_f) and maximum absolute value of r across strata ($Max(|r|$). *Correlation coefficient of 0.31 for elevation was only observed in one stratum.

| Covariate | r_f | $Max(r)$ |
|-----------------------------------|-------|------------|
| Elevation | -0.21 | 0.31* |
| Grazing intensity index | 0.64 | 0.19 |
| log annual precipitation | -0.51 | 0.25 |
| log catchment area | 0.51 | 0.23 |
| log Cl^- | 0.64 | 0.12 |
| log HCO_3^- | 0.56 | 0.21 |
| log $SO_4^{=}$ | 0.64 | 0.07 |
| Longitude | 0.63 | 0.24 |
| Percent catchment agriculture | 0.61 | 0.29 |
| Percent catchment urban | 0.34 | 0.12 |
| Percent open canopy | 0.49 | 0.11 |
| Percent substrate sand/fines | 0.65 | 0.26 |
| Riparian agricultural disturbance | 0.53 | 0.21 |
| Stream temperature | 0.48 | 0.20 |

TN co-varied strongly with a number of variables in this dataset. In the full dataset, TN was strongly positively correlated ($r > 0.5$) with grazing intensity, catchment area, Cl^- , HCO_3^- , SO_4^- , agriculture land use in the catchment and in the riparian zone, and percent substrate sand/fines (column labeled r_f in Table 2). It was also strongly negatively correlated with annual precipitation. After stratifying by propensity score, the strength with which each covariate was correlated with TN decreased across all strata. The maximum absolute value of correlation coefficients across all strata decreased to less than 0.3 for all variables except for elevation. Elevation was correlated with TN with $r = 0.31$ in only one stratum, and in all other strata, correlation strength was less than 0.21. Overall, stratifying by propensity score effectively reduced the degree to which other variables were correlated with TN.

Since other environmental variables are only weakly correlated with TN within each stratum, these variables are much less likely to confound regression estimates of the effects of TN on total richness calculated for each stratum. Thus, one can more confidently infer that regression estimates of TN effects reflect stressor-response relationships. The effects of TN on total invertebrate richness (Figure 35) varied across different strata. In Strata 1-3, increased TN had weakly positive (and not statistically significant) effects on total richness, whereas in Strata 4-6, increased TN was associated

Empirical Approaches for Nutrient Criteria Derivation
SAB Review Draft

with significant decreases in total richness. So, one could conclude from this analysis that in certain types of streams, increased TN concentration likely caused decreases in total richness. In Strata 4-6, a one unit change in log(TN) was associated with a loss of 6 to 9 taxa.

Propensity score analysis can be applied to any field dataset, given enough covariate data, and thus, can be used to estimate the effects of nutrients on the chosen response variable. Because the effects of covariates are greatly curtailed by this analysis approach, the estimated effects can be more confidently attributed to stressor-response relationships.

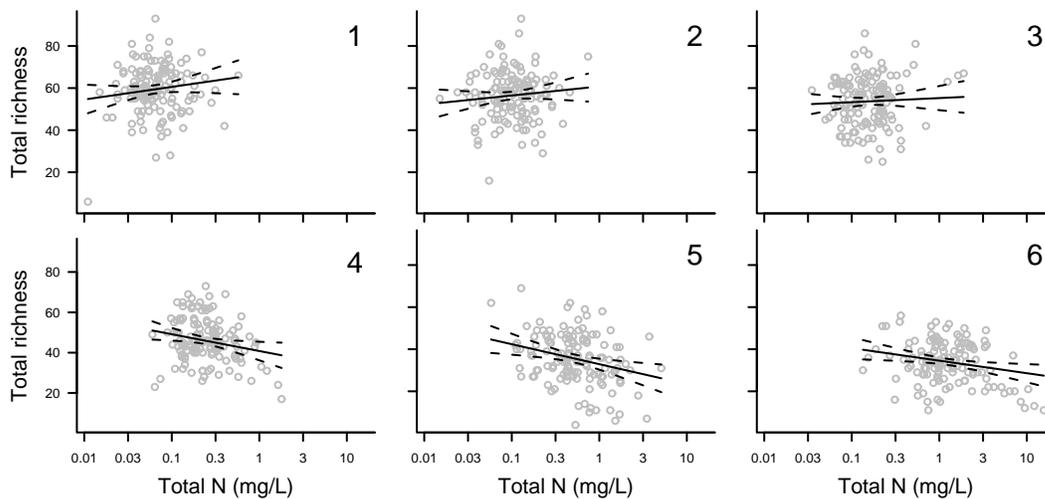


Figure 35. Relationships between TN and total invertebrate richness within each stratum. Solid lines are the linear regression estimates, and dashed lines are the 95% confidence intervals around those estimates. Strata roughly correspond to a gradient of human disturbance, from well-shaded, forested streams (Stratum 1) to open canopy streams in grazed areas (Stratum 6).

Appendix C – Conditional Probability Analysis

Conditional probability is fundamental to all of probability and statistics. It is generally covered in most probability texts and has been used in the analysis of data from numerous disciplines (e.g., Henry 1969, Mitchell 1991, Ming Tan 1996, Greenland and Bonow 2008). It is fundamental in Bayesian statistics where one views probability as a degree of belief (or confidence): how much confidence does one have in a specific result if certain information is true (Ashby 2006)? The application of conditional probability in the analysis of thresholds in environmental data is quite recent. This appendix describes application of conditional probability in the analysis of environmental information for exploring the relationship between candidate criterion values and the probability that a selected response variable will exceed a threshold indicating that a waterbody is not meeting its designated use. This relationship can further be analyzed to help identify candidate criteria. The approach is referred to here as conditional probability analysis (CPA), and most of the material in this section is drawn from Paul and McDonald (2005).

Description and Data Requirements

The probability of an event y is denoted as $P(y)$. A conditional probability is the probability of an event y when it is known that some other event x^* has occurred, and is denoted as $P(y|x^*)$. An example of a conditional probability would be the probability of degraded macroinvertebrate condition in a stream, given that lead concentrations in sediments exceed $10 \mu\text{g/L}$.

For use in developing and evaluating candidate criteria, conditional probabilities can be expressed in terms of the probability of not meeting the designated use, given that the observed stressor values exceed a candidate criterion. This probability can be written as $P(y = 1 | x \geq x_c)$, where $y = 1$ denotes not meeting a designated use (unacceptable conditions), $y = 0$ denotes meeting designated use (acceptable conditions), x is the nutrient (stressor variable), and x_c is the candidate criterion. For example, one could calculate the probability of a stream not meeting its designated use, given that the total phosphorus concentration in the stream exceeds $120 \mu\text{g/L}$. For discussion here, assume that as the magnitude of the stressor increases, the probability of not meeting the designated use also increases.

The observed stressor values, x_i , are treated as random values. If the data were acquired with a probability design (Cochran 1977), then the inclusion probability (IP) for a sample provides the probability that the sample was selected for a given draw from the target population. IPs are used as weighting factors for statistical estimation from the data. For a probability sample, every sample element in the target population has a non-zero probability of being selected. For non-probability samples, all samples are analyzed as having equal weights.

The response variable, y_i , is a random variable, either from observations or from a modeled stochastic function. Each value of y_i is paired with an observed value of the stressor, x_i . Then conditional probabilities can be calculated using the following relationship (Hogg and Ledolter 1992),

$$P(y = 1 | x \geq x_c) = \frac{P(y = 1, x \geq x_c)}{P(x \geq x_c)}$$

where $P(y = 1, x \geq x_c)$ is the joint probability, that is, both the stressor exceeds the candidate criterion value and the waterbody does not meet its designated use.

The following two-step procedure is used to estimate the conditional probability not meeting a designated use, given that the observed stressor value is exceeded in a particular dataset:

1. Identify a subset of the samples for which $x \geq x_c$.
2. For the subset of the samples identified at step 1, estimate the probability of not meeting the designated use.

This two-step procedure can be repeated for values of x_c over the entire range of observed x_i , and the resulting numbers describe an empirical conditional probability curve. The response variable is either from observations paired with the stressor variable or modeled as a stochastic function of the stressor variable. If the response variable is from observations, confidence intervals (CIs) for the empirical curve are estimated by one of two methods: (1) bootstrap resampling of the original data pairs (Manly 1997) or (2) assume that the subset of values for each x_c value can be treated as a weighted simple random sample (SRS).

Thresholds in the empirical conditional probability curve, if they exist, are identified by various techniques that can include changepoint analysis. A changepoint is defined as the value x_R ($1 \leq R < N$) that divides the response (empirical conditional probability curve) into 2 groups, $y_1 \dots y_{R-1}$ and $y_R \dots y_N$ each with distinct characteristics defined in some manner (Qian et al. 2003). The corresponding value of the stressor, x_R , is called the threshold. Changepoint analysis is a well-established procedure in the literature (Barry and Hartigan 1993; Groger et al. 2007).

The techniques for the identification of thresholds include: deviance reduction (see section 3.3.1), non-overlapping confidence intervals, and change in curvature in fit of a non-linear double logistic curve. These techniques are briefly described and illustrated in the examples that follow. It should be noted that all thresholds should be evaluated for significance to establish that the estimated threshold represents a real change in the characteristics of the response variable and is not just an artifact of the method.

To apply CPA to a dataset for criteria development, at least 3 conditions should be satisfied. They are:

1. The data should be sense representative of the aquatic resource for which criteria are to be derived so that results can be applied to unsampled locations. Data should be unbiased (e.g., for what purpose were the data originally collected?), and ideally, collected using a probability-based design. However, from a practical point of view, other data might be acceptable as long as it can be shown that the distribution of values is similar to the distribution of data collected using a probability-based design.
2. A metric or measurement must quantify the levels of the stressor of interest. If the metric is a surrogate for the stressor of interest, then its relationship to the stressor of interest should be clearly articulated.
3. A threshold in the response is known (as described in step 1) that quantitatively divides the aquatic resource into groups meeting and not meeting the designated use.

For use of conditional probability analysis by those unfamiliar with statistical calculations, a spreadsheet module has been developed and is available for conducting CPA (Hollister et al. 2008). While the calculations are done in R, the user interface is an Excel add-in. It is available at <http://www.epa.gov/emap/nca/html/regions/cprob/> (Figure 36). An example of the use of CPA for identification of candidates in stressor identification is available at www.epa.gov/caddis.

CProb 1.0: A Tool for Conditional Probability Analysis

CProb 1.0 is a Microsoft Excel Add-in, developed with the R language and environment for statistical computing, the R(D)Com Server and Visual Basic for applications, intended to aid in conditional probability analysis. The tool uses Microsoft Excel as a front-end for the R language and calculates conditional probability and bootstrapped estimates of confidence intervals. Additionally, optional scatterplots, cumulative distribution functions, and conditional probability plots may be generated.

Download Instructions:

1. Download [CProb_v1.zip](#) to a local folder
2. Open the zip file
3. Open CProb_ReadMe.pdf and follow the install instructions.

Citation:

If your use of CProb 1.0 results in a peer-reviewed publication or other scientific report, please use the following citation:

Hollister, J.W., H.A. Walker, J.F. Paul. (2008) CProb: A Computational Tool for Conducting Conditional Probability Analysis. *Journal of Environmental Quality*. 37 (6):2392-2396. [Abstract](#).

Questions or concerns?

Contact Jeff Hollister at hollister.jeff@epa.gov

Figure 36. CPROB add-in module for conducting conditional probability analysis.

Example Applications

In the two datasets used in these example applications, the site selections were made with a probability-based design. For all of the examples presented here, the sample values are treated as having equal weight to simplify the examples. The results of these examples would be applicable only to the sites that were sampled by the study. Because of this simplification, insights derived from these examples may not be generally applicable to the study area.

The first example uses data from the EMAP Northeast Lakes Survey (Appendix A). Chlorophyll *a* and total phosphorus water samples are used for the identification of candidate criterion for phosphorus with chlorophyll *a* as the response. A scatter plot of chlorophyll *a* vs. total phosphorus (log-log transformed) (Figure 12) indicates that the chosen stressor and response variables are strongly associated with one another.

The cumulative distribution function (CDF) of total phosphorus for all of the lakes is shown in Figure 37. Also shown in this figure are conditional CDFs (CCDFs) for eutrophic lakes and non-eutrophic lakes, where chlorophyll *a* is used as demarcation (threshold) between eutrophic and non-eutrophic. For this exercise, chlorophyll *a* = 15 µg/L is the

assumed threshold for the response variable that determines when these lakes do not meet their designated use. The CCDFs indicate a clear distinction between the distributions in these two groups.

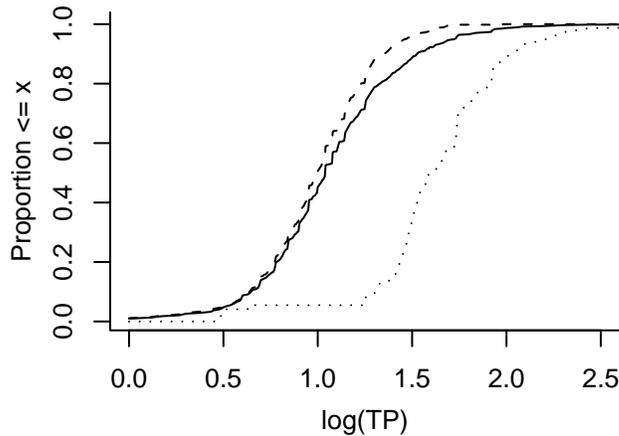


Figure 37. Cumulative distribution function (CDF) and conditional cumulative distribution functions (CCDFs) for total phosphorus concentrations for the EMAP Northeast Lakes Survey data. Dotted line: conditioned on chlorophyll $a < 15$, dashed line: conditioned on chlorophyll $a > 15$. Units are $\mu\text{g/L}$ of $\log(\text{TP})$.

The empirical conditional probability curve for the probability of eutrophic conditions ($\geq 15 \mu\text{g/L}$ chlorophyll a , using observed responses) given that the observed phosphorus exceeds the value on the x-axis is shown in Figure 38. The curve indicates that there is a 100% probability of observing a lake in eutrophic condition given that total phosphorus level in the lake exceeds $50 \mu\text{g/L}$. Possible thresholds from visual observation of the curve occur around 25 and $50 \mu\text{g/L}$ for phosphorus, depending if one incorporates the confidence limits or not. The threshold from non-overlapping confidence limits is identified by drawing a horizontal line from the upper limit of the unconditional probability value (value at extreme left) and determining where it crosses the lower limit of the empirical curve, which is at $10 \mu\text{g/L}$.

The non-overlapping confidence limit threshold ($10 \mu\text{g/L}$ total phosphorus) is the lower limit for a criterion of what could be detected as statistically different from what would be expected if observations were made across lakes in the whole geographic area (the unconditional probability). An upper limit for a criterion would be for 100% probability of a lake being eutrophic ($50 \mu\text{g/L}$ total phosphorus).

Also shown on Figure 38 (red line) is a non-linear least-squares curve fit to the double logistic curve:

$$P(y = 1 | x \leq x_0) = \begin{cases} 1 + \frac{D-1}{1 + \exp(B_1(x_c - x_0))}, & \text{for } x_c \leq x_0 \\ 1 + \frac{D-1}{1 + (B_2(x_c - x_0))}, & \text{for } x_c > x_0 \end{cases}$$

Where D , B_1 , B_2 , and x_0 are the regression parameters, with x_0 as the value where the change in curvature (B_1 to B_2) occurs. The change of curvature is at 21.1 $\mu\text{g/L}$ total phosphorus, with 95% confidence limits of 20.1 – 22.6. The 95% confidence limits of the slopes do not overlap, so this change point is significant. The deviance reduction changepoint is 20 $\mu\text{g/L}$ total phosphorus, with 95% confidence limits of 18-23 $\mu\text{g/L}$.

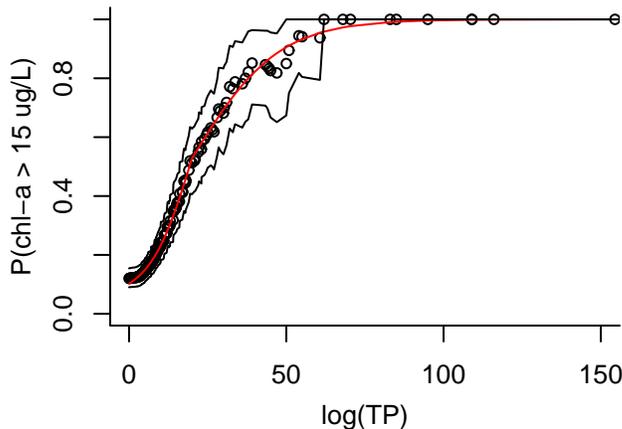


Figure 38. The empirical conditional probability curve (black open circles) for observing a eutrophic lake (chlorophyll $a > 15 \mu\text{g/L}$) if the total phosphorus concentration in the lake exceeds the x-axis value. Solid lines are 95% confidence limits. Red solid line is non-linear curve fit of double logistic curve. Units are $\mu\text{g/L}$ of $\log(\text{TP})$.

In summary, the results of the first example using CPA with non-overlapping confidence intervals for the EMAP Northeast Lakes Survey data with a use attainment threshold of 15 $\mu\text{g/L}$ chlorophyll a and equal weighting to all of the values, indicate a minimum candidate changepoint of 10 $\mu\text{g/L}$ and a maximum of 50, while non-linear curve fit determined changepoint at 21.1 and deviance reduction determined a candidate of 20 $\mu\text{g/L}$.

The second example uses data from EMAP West Xeric region streams (Appendix A). The response variable is total species macroinvertebrate richness and the nutrient variable is total nitrogen. The threshold for total species macroinvertebrate richness separating acceptable from unacceptable conditions was assumed to be 40 for this exercise.

The cumulative distribution function (CDF) of total nitrogen for all of the sampled streams is shown in Figure 39. Also shown in this figure are conditional CDFs (CCDFs) for streams with less than 40 total species richness and those with greater than or equal to 40. The distributions of nitrogen values differ across these two groups.

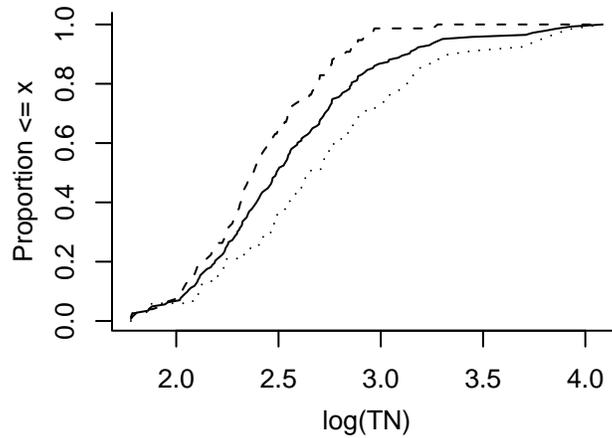


Figure 39. Cumulative distribution function (CDF) and conditional cumulative distribution functions (CCDFs) for EMAP-West Xeric region streams. Solid line: CDF for log(TN), dashed line: CCDF for log(TN) conditioned on total species richness ≥ 40 , and dotted line: CCDF for log(TN) conditioned on total species richness < 40 . Units are $\mu\text{g/L}$ of log(TN).

The empirical conditional probability curve for the probability of total richness < 40 given that the observed total N exceeds the value on the x-axis is shown in Figure 40. The curve indicates that there is a 100% probability of observing a richness less than 40 when $\log(\text{TN}) > 3.3$. The confidence limits (95%) determined by bootstrap resampling are shown as solid lines. The threshold from non-overlapping confidence limits is identified by drawing a horizontal line from the upper limit of the unconditional probability value (value at extreme left) and determining where it crosses the lower limit of the empirical curve, which is at approximately $\log(\text{TN}) = 2.7$.

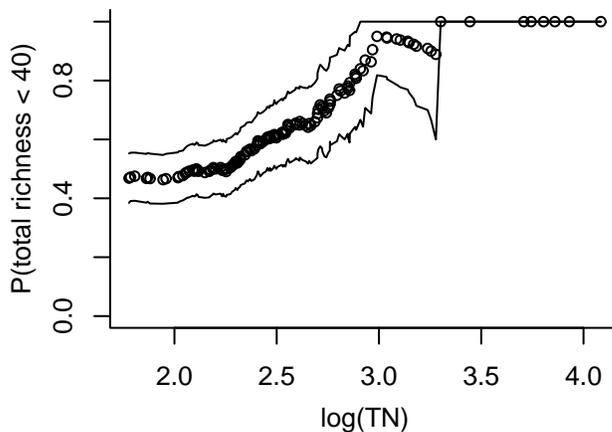


Figure 40. The empirical conditional probability curve (black open circles) for observing a stream with total richness < 40 if the total nitrogen concentration in the stream exceeds the x-axis value. Solid lines are bootstrap 95% confidence limits. Units are $\mu\text{g/L}$ of log(TN).

Appendix D – Acronyms

CART – Classification and Regression Trees
CDF – Cumulative Distribution Function
Chl *a* – Chlorophyll *a*
CPA – Conditional Probability Analysis
CWA – Clean Water Act
EDA – Exploratory data analysis
EMAP – USEPA Environmental Monitoring and Assessment Program
IBI – Index of Biological Integrity
MLR – Multiple linear regression
nCPA – Non-parametric changepoint analysis
ORD – Office of Research and Development
OST – Office of Science and Technology
Q-Q – Quantile-quantile
QR – Quantile Regression
RMSPE – Root-mean squared prediction error
SEM – Structural Equation Models
SLR – Simple Linear Regression
TN – Total Nitrogen
TP – Total Phosphorus
USEPA – United States Environmental Protection Agency
WQS – Water Quality Standards

Appendix E – Glossary

algae

Chiefly aquatic, eukaryotic one-celled or multicellular plants without true stems, roots and leaves that are typically autotrophic, photosynthetic, and contain chlorophyll.

algal biomass

The weight of living algal material in a unit area at a given time.

aquatic life use

A beneficial use designation in which the waterbody provides suitable habitat for survival and reproduction of desirable fish, shellfish, and other aquatic organisms.

assemblage

An association of interacting populations of organisms in a given waterbody. Examples of assemblages used for biological assessments include : algae, amphibians, birds, fish, herps (reptiles and amphibians), macroinvertebrates (insects, crayfish, clams, snails, etc.), and vascular plants.

benthos/benthic

The assemblage of organisms associated with the bottom, or the solid-liquid interface of the aquatic system. Generally applied to organisms in the substrata.

biological assessments or bioassessments

Evaluation of the biological condition of a waterbody using biological surveys and other direct measurements of resident biota in surface waters.

biological criteria or biocriteria

Narrative or numeric expressions that describe the biological condition (structure and function) of aquatic communities inhabiting waters of a designated aquatic life use. Biocriteria are based on the numbers and kinds of organisms present and are regulatory-based biological measurements.

biological integrity

The ability of an aquatic ecosystem to support and maintain a balanced, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of natural habitats within a region.

biological monitoring or biomonitoring

Use of a biological entity as a detector and its response as a measure to determine environmental conditions. Toxicity tests and ambient biological surveys are common biological monitoring methods.

cause

That which produces an effect (a general definition).

chlorophyll a

A complex molecule composed of four carbon-nitrogen rings surrounding a magnesium atom; constitutes the major pigment in most algae and other photosynthetic organisms; is used as a reliable index of algal biomass.

confidence interval

An interval defined by two values, called confidence limits, calculated from sample data using a procedure which ensures that the unknown true value of the quantity of interest falls between such calculated values in a specified percentage of samples. Commonly, the specified percentage is 95%; the resulting confidence interval is then called a 95% confidence interval. A one-sided confidence interval is defined by a single calculated value called an upper (or lower) confidence limit.

cumulative distribution:

A means of representing the variation of some attribute by giving running totals of the resource with attribute values less than or equal to a specified series of values. For example, a cumulative areal distribution of lakes would give, for any value of area, the total area covered by lakes with individual area less than or equal to alpha. A cumulative frequency distribution for lake area would give the total number of lakes with area less than or equal to alpha. The cumulative distribution function (cdf) of some specified attribute of a population is the function $F(x)$ that gives the proportion of the population with value of the attribute less than or equal to x , for any choice of x . For example, if the attribute was lake area in hectares, $F(a)$ would give the proportion of lakes with area less than or equal to a ha. (In some cases, the word "cumulative" may be omitted in discussions of the cdf, and the cdf is called the distribution function.)

criteria

Elements of State water quality standards, expressed as constituent concentrations, levels, or narrative statements, representing a quality of water that supports a particular use. When criteria are met, water quality will generally protect the designated use.

designated uses

Uses defined in water quality standards for each waterbody or segment whether or not the use is being attained.

diatom

Microscopic algae with cell walls made of silicon and have two separating halves.

ecoregion

A region defined by similarity of climate, landform, soil, potential natural vegetation, hydrology, and other ecologically relevant variables.

empirical

Relying upon or gained from experiment or observation.

endpoint

An observable or measurable biological event or chemical concentration (e.g., metabolite concentration in a target tissue) used as an index of an effect of a chemical exposure.

eutrophic

Abundant in nutrients and having high rates of productivity frequently resulting in oxygen depletion below the surface layer.

eutrophication

The increase of nutrients in [waterbodies] either naturally or artificially by pollution.

gradient

The rate of change in value of a physical or chemical parameter per unit change in position.

habitat

A place where the physical and biological elements of ecosystems provide a suitable environment including the food, cover, and space resources needed for plant and animal livelihood.

impairment

Detrimental effect on the biological integrity of a waterbody caused by an impact that prevents attainment of the designated use.

linear model

Linear models have the form,

$$f(x)=\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3\dots$$

in which for any number of independent or explanatory variables (i.e., $x_1, x_2, x_3 \dots$), each independent or explanatory variable in the model is multiplied by an unknown parameter (i.e., $\beta_1, \beta_2, \beta_3 \dots$), there is at most one unknown parameter with no corresponding independent or explanatory variable (i.e., β_0), and all of the individual terms are summed to produce the final function model.

Although such a function may not describe a straight line, it is said to be linear in the parameters, because the problem can be reduced to system (i.e., one to many) of algebraic (i.e., linear) equations that can be solved for unique values of the unknown parameters (i.e., $\beta_0, \beta_1, \beta_2, \beta_3 \dots$).

macroinvertebrates

Animals without backbones of a size large enough to be seen by the unaided eye and which can be retained by a U.S. Standard No. 30 sieve (28 meshes per inch, 0.595 mm openings).

metric

A calculated term or enumeration representing some aspect of biological assemblage, function, or other measurable aspect and is a characteristic of the biota that changes in some predictable way with increased human influence. A multimetric approach involves combinations of metrics to provide an integrative assessment of the status of aquatic resources.

mesotrophic

Having a nutrient loading resulting in moderate productivity.

µg/L

micrograms per liter, 10^{-6} grams per liter

mg/L

milligrams per liter, 10^{-3} grams per liter

minimally impaired

Sites or conditions with slight anthropogenic perturbation relative to the overall region of the study.

model

A mathematical function with parameters that can be adjusted so the function closely describes a set of empirical data.

monitoring

Periodic or continuous surveillance or testing to determine the level of compliance with statutory requirements and/or pollutant levels in various media [air, soil, water] or in humans, plants, and animals.

multimetric

Analysis techniques using several measurable characteristics of a biological assemblage.

multivariate

Type of statistics that relates one or more independent (explanatory) variables with multiple dependent (response) variables.

nutrients

Elements (*e.g.*, nitrogen and phosphorus) essential for the growth of organisms.

oligotrophic

Trophic status of a waterbody characterized by a small supply of nutrients (low nutrient release from sediments), low production of organic matter, low rates of decomposition, oxidizing hypolimnetic condition (high DO).

periphyton

Associated aquatic organisms attached or clinging to stems and leaves of rooted plants or other surfaces projecting above the bottom of a waterbody.

prediction interval

A type of statistical interval that has a specified probability (commonly 95%) of enclosing the value of a "future" unit that is predicted based on the available data and does not belong to the sample used to generate the prediction, assuming that the future unit is drawn from the same population. In the context of regression analysis, standard prediction intervals relate to uncertainty when the fitted regression is used to predict the response (Y) variable for specific values of X variables. In cases of linear regression with a single X variable, prediction intervals associated with all X values are conventionally depicted as a band enclosing the fitted regression line, bounded by two curves that diverge as X increases in distance from the mean X, in either direction. Prediction intervals address both unit variation (e.g., as evaluated using sample quantiles) and statistical error in estimating unknown population parameters (e.g., in estimating a regression slope and intercept) and therefore can be distinguished from confidence intervals which only address the statistical error in parameter estimation.

precision

The degree to which replicate measurements of the same attribute agree or are exact.

quantile

The value of an attribute indexing a specified proportion of a population distribution or distribution function. Quartiles (25th, 50th, and 75th percentiles), the median (50th percentile), and other percentiles are special cases of quantiles.

random sampling

Generic type of probability sampling, randomness can enter at any stage of the sampling process.

reference conditions

Describe the characteristics of waterbody segments least impaired by human activities. As such, reference conditions can be used to describe attainable biological or habitat conditions for waterbody segments with common watershed/catchment characteristics within defined geographical regions.

reference site

Specific locality on a waterbody which is unimpaired or minimally impaired and is representative of the expected biological integrity of other localities on the same waterbody or nearby waterbodies.

resampling

Any of a variety of methods used for estimating the precision of sample statistics (mean, variances, or percentiles) using subsets (jackknifing) or random sampling with replacement (bootstrapping) of available data.

statistical significance

The probability that a result is not likely to be due to chance alone. By convention, a difference between two groups is usually considered statistically significant if chance could explain it only 5% of the time or less. Study design considerations may influence the a priori choice of a different level of statistical significance.

stratification, stratified random sampling

Type of probability sampling where a target population is divided into relatively homogenous groups or classes (strata) prior to sampling based on factors that influence variability in that population. In stratified sampling, a heterogenous environment is divided into homogenous strata or parts. Analysis of variance can be used to identify statistically different parameter means among the sampling strata or classes. The strata are the analysis of variance treatments.

stressor

Any physical, chemical, or biological entity that can induce an adverse response

stressor-response relationship

The relationship between the intensity, frequency, or duration of exposure to a stressor and the intensity or frequency of a biological response and/or a model of that relationship.

trophic state

The degree of nutrient enrichment of a waterbody.

turbidity

Cloudiness or opaqueness of a suspension. In this context, refers to the amount of suspended matter in the water column, usually measured in nephelometric turbidity units.

variable

A measurable attribute that can be used to evaluate or predict the condition of a waterbody.

variance

A measure of the variability or precision of a set of observations.

watershed

The area of land that drains water, sediment, and dissolved materials to a common outlet at some point along a stream channel. In American usage, *watershed* is synonymous with the terms *drainage basin* and *catchment*.