

2
2006

This is to certify that the
dissertation entitled

INFERRING REFERENCE CONDITIONS TO ASSESS THE
BIOLOGICAL INTEGRITY OF STREAMS AND RIVERS

presented by

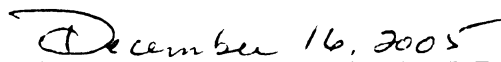
SCOTT LYLE ROLLINS

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Zoology



Major Professor's Signature



Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
JUL 06 2010		
JUL 02 10		

INFERRING REFERENCE CONDITIONS TO ASSESS THE BIOLOGICAL
INTEGRITY OF STREAMS AND RIVERS

By

Scott Lyle Rollins

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Zoology

2005

ABSTRACT

INFERRING REFERENCE CONDITIONS TO ASSESS THE BIOLOGICAL INTEGRITY OF STREAMS AND RIVERS

By

Scott Lyle Rollins

Inference models are often necessary to quantify the effects of human disturbance on aquatic ecosystems and to assess the current biological integrity of these systems. In my dissertation, I apply modern computational modeling approaches to to infer expected conditions in streams and rivers if human disturbance were minimized.

Autecological characteristics of diatom species can be used to infer environmental conditions when they cannot be measured directly. I evaluated standard methods and applied new methods to model diatom species responses to total nitrogen (TN), total phosphorus (TP), and pH, and to infer environmental conditions from diatom species composition. Current methods were relatively effective at describing species distributions along resource gradients, but failed to adequately reflect species responses to pH. Modern methods did not significantly improve standard inference models for any of the environmental variables.

Linear discriminant analysis (LDA) has been used to predict the expected taxonomic composition in the absence of human disturbance to assess biological integrity.

However, many biological responses to ecological gradients are nonlinear. A hybrid approach combining LDA with nonlinear predictions was applied using a Bayesian methodology. Diatom assemblages were used to classify minimally-disturbed sites throughout the western United States. Predictive models for the classes were then developed using LDA, recursive partitioning, and the new hybrid method. The hybrid method outperformed both LDA and recursive partitioning, suggesting that nonlinear determinants of diatom assemblages are important and that the hybrid method will improve our ability to assess the biological integrity of streams.

In the final chapter, I apply Bayesian statistical methods to inform the development of nutrient water quality criteria. The United States Environmental Protection Agency suggests using prior research, reference-based approaches, and stressor-response relationships to develop regulatory levels for nutrients in streams and rivers; however, a framework for integrating sources of information and endpoints is lacking. I provide a framework that integrates this information, explicitly acknowledges model uncertainty, and is easily communicated using measures of relative risk.

Dedicated to my wife, Krista.

ACKNOWLEDGEMENTS

I thank my dissertation advisor, R. J. Stevenson, and my dissertation guidance committee, S. K. Hamilton, R. W. Merritt, and G. G. Mittelbach for helpful comments regarding the work in my dissertation, as well as other research that is not included here. In addition to comments provided by my committee, discussions with Y. Pan, H. Leland, and D. Charles at the 50th Annual Meeting of the North American Benthological Society improved the second chapter. Discussions with S. Qian and Y. Pan, as well as comments from B. Bixby and an anonymous presentation judge at the 53rd Annual Meeting of the North American Benthological Society improved the third chapter. The fourth chapter evolved out of several meetings and discussions with members of the Michigan Department of Environmental Quality, particularly S. Heaton, S. Holden, and S. Wolf, as well as K. Spence Cheruvellil and P. Soranno. The third chapter was made possible by many people who worked sun-up to sun-down for months with few days off. These people included A. Andersen, E. Billman, Y. Cao, B. Creutzburg, K. Duester, M. Helfmeyer, S. Hockett, A. May, C. Mellon, J. Olson, J. Ostermiller, and K. Rollins. GIS data resulted from the work of J. Olson, R. Hill, N. Detenbeck, and C. Riseng. Designated use

attainment data for the State of Michigan was provided in digital format by J. Wuycheck. K. Manoylov taught me the fundamentals of diatom taxonomy and provided an incredible model of perseverance. Several people assisted with field research at the University of Michigan Biological Station that was not included in my dissertation, including J. Heinlein, C. Parker, K. Rollins, and S. Wolf. Members of the Stevenson lab and Merritt lab, as well as many students in the EEBB program will be lifelong friends. I am particularly grateful for the friendship of Sarah Wolf. Most of all, I am thankful for the support of my wife, Krista. I couldn't have done this without her.

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
CHAPTER ONE	
INTRODUCTION.....	1
References.....	9
CHAPTER TWO	
DIATOM SPECIES' RESPONSES TO STRESSORS AND RESOURCES: ARE THEY GAUSSIAN AND DOES IT INFLUENCE ENVIRONMENTAL INFERENCE MODELS?.....	12
Introduction.....	13
Methods.....	20
Data.....	20
Species Response Models and Evaluation.....	22
Inference Models.....	26
Results.....	31
Evaluation of the Gaussian Model.....	31
Comparison of TP, TN, and pH Failure Rates....	31
Performance of Inference Models.....	32
Discussion.....	40
References.....	46
CHAPTER THREE	
PREDICTING DIATOM ASSEMBLAGES IN MINIMALLY DISTURBED STREAMS USING A NEW HYBRID MODELLING APPROACH.....	49
Introduction.....	50
Methods.....	56
Data.....	56
Clustering Sites by Taxonomic Composition.....	60
Predicting Class Membership.....	61
Results.....	66
Clustering of Sites.....	66
LDA Model Summary.....	66
CART Model Summary.....	74
Model Performance.....	77
Discussion.....	83
References.....	88

CHAPTER FOUR

A RELATIVE RISK FRAMEWORK FOR DEVELOPING NUTRIENT WATER QUALITY CRITERIA BY INFERRING REFERENCE CONDITIONS AND

THE PROBABILITY OF LOSING VALUED ECOLOGICAL ATTRIBUTES..	92
Introduction.....	93
Methods.....	101
Environmental Thresholds.....	101
Predictive Models.....	105
Defining Reference Conditions.....	107
Establishing Benchmarks Using Relative Risk...	108
Results.....	115
Discussion.....	125
References.....	131

LIST OF TABLES

Table 1. Comparison of inference models using GLM and GAM derived parameters. A 5% increase in correlation and 10% decrease in RMSE was considered a significant improvement by using the GAM inference model.....	39
Table 2. Self-organizing map classification summary, including the number of sites place in each class (N) and the associated mean squared error of each class (MSE).....	67
Table 3. Proportion of sites within each class in which diatom genera were observed.....	68
Table 4. Performance of linear discriminant analysis (LDA), recursive partitioning (CART), and the Bayesian hybrid predictive classification models with respect to misclassification rates, the odds of correctly classifying a site relative to the null model (random assignment to a class) and confidence-weighted ratios (CWR) of correct-to-incorrect classifications.....	78
Table 5. Observed and predicted classifications using linear discriminant analysis (LDA).....	79
Table 6. Observed and predicted classifications using the classification tree (CART).....	80
Table 7. Observed and posterior predicted classifications using the hybrid method combining linear discriminant analysis (LDA) and classification tree (CART) predictions.....	82
Table 8. Posterior estimation of parameters in the predictive model used to infer log-transformed total phosphorus (mg/L), $\log(\text{TP}) = \beta_0 + \beta_1[\ln(\text{Channel Length})] + \beta_2[\ln(\text{Precipitation})] + \beta_3[\% \text{ Urban}] + \beta_4[\% \text{ Crop}] + \varepsilon$	122

LIST OF FIGURES

Figure 1. Gaussian curve commonly used to model species responses along environmental gradients. The species' optimum is the value of the environmental variable at which the probability of occurrence is maximized ($\hat{\theta}$). Dashed lines indicate the tolerance range ($\hat{\theta}-\hat{\sigma}$ to $\hat{\theta}+\hat{\sigma}$). The tolerance value is usually indicated by the estimated standard deviation ($\hat{\sigma}$)	14
Figure 2. Stream sites sampled in the Mid-Atlantic Highlands region of the eastern United States as part of the United States Environmental Protection Agency's Environmental Monitoring and Assessment Program.....	21
Figure 3. Example diatom species response curves. Solid lines represent the Gaussian response curve estimated by GLM. Dashed lines represent GAM response curves and associated 95% confidence intervals. (a) Example of a GLM that failed because it was not contained within the 95% confidence interval of the GAM. (b) Example of a GLM that failed because it was U-shaped and because it was not contained within the 95% confidence interval of the GAM. (c) Example of a GLM that failed because the predicted optimum is outside of the observed range of the environmental variable. (d) Example of a GLM model that passed all evaluation criteria.....	24
Figure 4. Examples of GLM (above) and GAM (below) response curves with estimated central borders (vertical dashed lines) from which species' tolerance ranges were calculated.....	28
Figure 5. Observed vs. diatom-inferred total phosphorus in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.....	33
Figure 6. Observed vs. diatom-inferred total phosphorus in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.....	34
Figure 7. Observed vs. diatom-inferred total nitrogen in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.....	35

Figure 8. Observed vs. diatom-inferred total nitrogen in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.....	36
Figure 9. Observed vs. diatom-inferred pH in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.....	37
Figure 10. Observed vs. diatom-inferred pH in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.....	38
Figure 11. Location of stream reaches sampled throughout the western United States.....	57
Figure 12. Classification tree (CART) for predicting diatom class membership using environmental variables. Decision rules at each split indicate which branch to take. Terminal nodes indicate predicted class membership. Decision rules that did not fit neatly onto the figure are indicated by circled numbers at the node and are as follows: 1) longitude ≥ -109.4 to the left, longitude < -109.4 to the right; 2) maximum temperature ≥ 13.23 to the left, maximum temperature < 13.23 to the right; 3) hydrologic stability ≥ 0.09905 to the left, hydrologic stability < 0.09905 to the right; 4) bedrock depth ≥ 46.54 to the left; bedrock depth < 46.54 to the right; 5) number of wet days < 98.7 to the left, number of wet days ≥ 98.7 to the right; 6) soil organic matter < 1.038 to the left, soil organic matter ≥ 1.038 to the right; 7) dominant geology granitic, mixed, sedimentary, or volcanic to the left, dominant geology gneiss or ultra-mafic to the right; 8) percent canopy cover ≥ 62.34 to the left, percent canopy cover < 62.34 to the right. Several variables are described in more detail in the text.....	75
Figure 13. Steps involved in developing candidate nutrient criteria using the relative risk framework presented in this document. Prior information boxes represent understanding of a given relationship prior to analyzing the data. Priors include an estimate of central tendency, as well as uncertainty associated with the prior information. Posterior inference boxes represent the results of Bayesian analysis that integrates prior information and data.....	100

Figure 14. Probability of not attaining Category 2 status from the Michigan Department of Environmental Quality 305(b) report as a function of percent urban (a and b) and percent crop (c) land use in the watershed. Curves represent LOESS fits (solid line) and pointwise 95% confidence intervals (dashed lines). Figure 1b is equivalent to figure 1a, with the exception of a log-transformed x-axis to improve resolution at the low end of the gradient.....	109
Figure 15. Graphical representation for calculating the probability that inferred reference nutrient concentrations are higher than the stressor-response threshold and the probability that the current nutrient concentration exceeds the stressor-response threshold (next page).....	113
Figure 16. Total phosphorus-water column chlorophyll threshold. Cumulative frequency distributions of the prior changepoint (dashed line) and posterior changepoint (solid line) densities indicate risk of exceeding the changepoint separating a state of low mean chlorophyll and high mean chlorophyll.....	116
Figure 17. Prior, likelihood, and posterior densities of natural-log-transformed chlorophyll ($\mu\text{g/L}$) below the total phosphorus threshold.....	117
Figure 18. Prior, likelihood, and posterior densities of natural-log-transformed chlorophyll ($\mu\text{g/L}$) above the total phosphorus threshold.....	118
Figure 19. Prior, likelihood, and posterior densities of the predictive model parameter for percent urban land use in the watershed.....	120
Figure 20. Prior, likelihood, and posterior densities of the predictive model parameter for percent crop land use in the watershed.....	121
Figure 21. Detailed presentation of the Cass River site analysis, showing posterior densities in predicted reference total phosphorus and the changepoint separating sites characterized by low mean chlorophyll from sites characterized by high mean chlorophyll. Current total phosphorus and the associated relative risk (RR) at that level is shown, along with the TP benchmark determined at a relative risk of one.....	124

CHAPTER ONE

INTRODUCTION

Ecology has a long history of applying mathematics in theoretical settings. Population growth models (Caswell 2001, Malthus 1798), competition models (Tilman 1982), and predator-prey models (Lotka 1925, Volterra 1926) are well known examples of theoretical applications of mathematics in ecology. Theoretical models provide a useful way to describe generalizable ecological processes or mechanisms and to explore patterns in a simplified and easily tractable system. Theoretical models, however, often operate under many assumptions that do not hold in natural systems, which can limit their ability to predict natural patterns. This difficulty scaling from simple to complex systems can burden the application of theoretical and experimental results to real-world problems (Clark 2005, Carpenter 1996, Levin 1992).

Phenomenological models are also commonly used in ecology. These empirical models describe patterns without explicitly representing the underlying mechanisms. The complex nature of ecosystems often necessitates the use of phenomenological approaches when informed decisions need to be made but mechanisms are not fully understood, data for some variables are not available, or data are at a different resolution than the mechanisms leading to patterns. Much of the data available for predicting the effects of environmental change on ecosystems have these

problems (Clark et al. 2001), making it difficult to apply process-based models. Therefore, phenomenological models are often applied. However, statistical models are sometimes selected for convenience without regard for theory or whether a model is appropriate for the patterns being described (Austin 1980, 1999). A pragmatic approach that combines theory and phenomenology is likely to be the most effective method for informing management decisions (Clark 2005, Carpenter 2002), yet data limitations are likely to limit the ability to effectively incorporate processes-based predictions (Clark et al. 2001).

Quantifying the effects of human activities on biological systems has recently received significant attention with respect to forecasting future ecosystem states (e.g., Clark 2003, Pielke and Conant 2003, Peterson et al. 2003, Clark et al. 2001). Policy makers and environmental managers also require retrospective ecological predictions and much of the discussion focused on forecasting applies to retrospective analysis. Bioassessment retrospectively examines the effects that anthropogenic activities have on the biological attributes of ecosystems, providing an appraisal of existing ecosystem damage. *Retrospective assessments* have been distinguished from *predictive assessments* that attempt to forecast the effects of management actions on ecosystems (Cairns and

Niederlehner 1995). However, determining whether a system is currently impaired requires inferring what it would look like, absent of the activities potentially causing impairment. These conditions are commonly called reference conditions in bioassessment. Predictive models provide a way to quantify reference conditions and their associated uncertainty. Predictive models facilitate both prospective and retrospective assessments by inferring how a system would look if anthropogenic activities differed from their current state.

Complex ecological models have become more accessible due to improved performance of desktop computers and advancements in computational statistics (Clark 2005). It is now possible to evaluate standard modeling approaches to determine whether they are consistent with observed patterns or inconsistent with theory. It is also possible to apply more complex predictive modeling approaches that account for uncertainty and/or incorporate processes-based components (Clark 2005). In my dissertation, I apply modern computational modeling approaches to evaluate current methods, improve predictions, and quantify uncertainty when inferring reference conditions for the retrospective assessment of stream and river health. In the second chapter, the assumptions of current models being used to describe diatom species' responses along

environment gradients and to infer historical environmental conditions are evaluated. These models assume that species respond in a Gaussian function (i.e., symmetrical and bell-shaped) to environmental variables such as pH. Parameters derived from these individual species models are used to infer environmental conditions when the environmental data cannot be effectively measured directly. One well known application of this method in paleolimnology is the historical reconstruction of lake chemistry (e.g., Birks et al. 1990).

Ecologists have argued that there is little theoretical justification for expecting universal Gaussian species' responses to all types of environmental gradients (Oksanen and Minchin 2002). Researchers have shown that Gaussian models often fail to effectively describe the presence of terrestrial plants along environmental gradients (Austin 1976, 1980, 1999, 2002; Oksanen and Minchin 2002). Despite this research, Gaussian models are continually applied to describe the presence of diatom species along environmental gradients of various types, with little acknowledgment of the assumptions involved, both statistically and with respect to ecological theory. Using generalized additive models (GAM), I evaluate whether Gaussian assumptions are appropriate for diatom responses to total phosphorus, total nitrogen, and pH in Mid-Atlantic Highlands streams.

Furthermore, I compare the ability of GAM inference models and Gaussian inference models to predict observed concentrations of phosphorus, nitrogen, and pH.

In the third chapter, a new method is presented to infer reference conditions for diatom assemblages using geology, hydrology, and climate variables. The current trend in the United States is to use minimally-disturbed or least-disturbed sites within a region to establish reference conditions for other sites within the region. If a site differs significantly from reference conditions, it indicates that the site has been affected by anthropogenic activities. Research suggests, however, that regionalizations are an ineffective way to classify sites (Hawkins et al. 2000), calling into question the usefulness of this method for establishing reference conditions. A different approach being applied in the United Kingdom and Australia creates stream classes using biological data, and then predicts class membership for new sites using geology, hydrology, and climate variables (Norris and Norris 1995, Clarke et al. 2003). The method used in these countries applies linear discriminant analysis (LDA), for which ecological data can be problematic (Williams 1983). A hybrid modelling approach that integrates linear (LDA) and non-linear (classification tree) model predictions is developed and applied to predict diatom genus composition

in minimally-disturbed reference streams throughout the western United States.

In the final chapter, statistical models are applied to aide in the development of site-specific nutrient criteria for streams and rivers in Michigan. Water quality criteria are established by States and tribes to protect attributes valued by the public. Developing nutrient criteria involves evaluating previous research, establishing expected conditions (i.e., reference conditions), and determining the effects of nutrients on valued ecological attributes (USEPA 2000). Expected conditions can be established using models that predict nutrient concentrations at minimal levels of human disturbance (Dodds and Oakes 2004). The relationship between nutrients and valued ecological attributes can be determined using a combination of previous research and quantitative patterns observed in available data (USEPA 2000). Thresholds, or large changes in valued ecological attributes over a narrow range of nutrients, are particularly useful for establishing water quality criteria because they can indicate benchmark nutrient concentrations that separate acceptable and unacceptable conditions in aquatic ecosystems (Stevenson et al. 2004, Qian et al. 2004, King and Richardson 2003). An effective method for integrating these steps is lacking. I describe a framework that

integrates these steps and summarizes the risk of exceeding environmental thresholds, relative to the risk of exceeding the threshold at lower levels of human disturbance. This framework uses Bayesian statistical models to infer expected levels of phosphorus when human disturbance is minimized and to quantify environmental thresholds. The framework is unique because 1) it explicitly integrates the findings of previous research, 2) it accounts for uncertainty, and 3) it reduces effects-based endpoints and inferred reference conditions to a single value, relative risk, which is easily communicated.

REFERENCES

- Austin, M. P. 1976. On non-linear species response models in ordination. *Vegetatio* 33: 33-41.
- Austin, M. P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11-21.
- Austin, M. P. 1999. A silent clash of paradigms: some inconsistencies in community ecology. *Oikos* 86: 170-178.
- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101-118.
- Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., and ter Braak, C. J. F. 1990. Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London B* 327: 263-278.
- Carpenter, S. R. 2002. Ecological futures: building an ecology for the long now. *Ecology* 83: 2069-2083.
- Carpenter, S. R. 1996. Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology* 77: 677-680.
- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd ed. Sinaur Associates, Inc. Publishers, Sunderland, MA.
- Clark, J. S. 2003. Uncertainty in ecological inference and forecasting. *Ecology* 84: 1349-1350.
- Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2-14.
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Jr., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., and Wear, D. 2001. Ecological forecasts: an emerging imperative. *Science* 293: 657-660.

- Clarke, R. T., Wright, J. F. and Furse, M. T. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling* 160: 219-233.
- Dodds, W. K., and Oakes, R. M. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. *Limnology and Oceanography Methods* 2: 333-341.
- Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28: 49-50.
- Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., and Stevenson, R. J. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society* 19: 541-556.
- King, R. S., and Richardson, C. J. 2003. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management* 31: 795-809.
- Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967.
- Lotka, A. J. 1925. *Elements of physical biology*. Williams and Wilkins, Baltimore, MD.
- Malthus, T. R. 1798. *An essay on the Principle of Population*.
- Norris, R. H., and Norris, K. R. 1995. The need for biological assessment of water quality: Australian perspective. *Australian Journal of Ecology* 20: 1-6.
- Oksanen, J., and Minchin, P. R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157: 119-129.
- Peterson, G. D., Carpenter, S. R., and Brock, W. A. 2003. Uncertainty and the management of multistate ecosystems: an apparently rational route to collapse. *Ecology* 84: 1403-1411.

- Pielke, R. A., and Conant, R. T. 2003. Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology* 84: 1351-1358.
- Qian, S. S., Pan, Y., and King, R. S. 2004. Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. *Ecological Indicators* 4: 29-37.
- Stevenson, R. J., Bailey, R. C., Harrass, M. C., Hawkins, C. P., Alba-Tercedor, J., Couch, C., Dyer, S., Fulk, F. A., Harrington, J. A., Hunsaker, C. T., and Johnson, R. K. 2004. Interpreting results of ecological assessments. Pages 85-111 in Barbour, M. T., Norton, S. B., Preston, H. R., and Thornton, K. W., eds. *Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making*. Society of Environmental Toxicology and Chemistry.
- Tilman, D. 1982. *Resource Competition and Community Structure*. Princeton University Press, Princeton, NJ.
- United States Environmental Protection Agency (USEPA). 2000. *Nutrient Criteria Technical Guidance Manual: Rivers and Streams*. Office of Water, Office of Science and Technology, U.S. Environmental Protection Agency. EPA-822-B-00-002.
- Volterra, V. 1926. Variations and fluctuations of the number of individuals in animal species living together. Pages 409-448 in Chapman, R. N., *Animal Ecology*. 1931. McGraw-Hill, New York, NY.
- Weinberg, W. 1908. Über den Nachweis der Verebung beim Menschen". *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 64: 368-382.

CHAPTER TWO

DIATOM SPECIES' RESPONSES TO STRESSORS AND RESOURCES:
ARE THEY GAUSSIAN AND DOES IT INFLUENCE ENVIRONMENTAL
INFERENCE MODELS?

INTRODUCTION

Diatom autecological data can be used to infer past environmental conditions (Charles and Smol 1988). These historic reconstructions can be used for environmental assessment to establish expected environmental conditions by estimating the state of an ecosystem prior to disturbance by human activities (Dixit et al. 1992, Battarbee and Charles 1987). Such inference models, or transfer functions, are commonly used to reconstruct historic environmental conditions in lakes and play a dominant role in the field of paleolimnology (Birks 1998).

One of the basic assumptions of these models is that the niche of a species can be adequately represented by a bell-shaped function. Niche theory, as generally depicted in text books (e.g., Ricklefs 1997, Krebs 1994) and applied in statistical models (e.g., Salden 1978, ter Braak and Looman 1986), suggests that species exhibit symmetrical, bell-shaped responses along environmental gradients (Figure 1). These "Gaussian" response curves describe the probability of occurrence, the proportional abundance, or density of a species as a function of some independent environmental variable. The value along the environmental gradient at which the function peaks is the maximum likelihood estimate and describes the species' optimum.

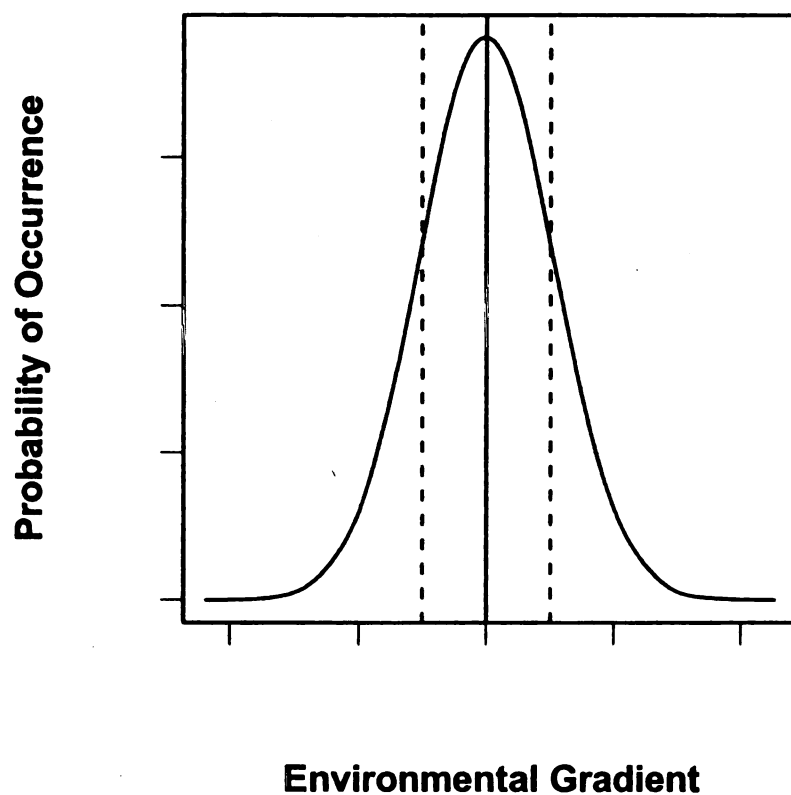


Figure 1. Gaussian curve commonly used to model species responses along environmental gradients. The species' optimum is the value of the environmental variable at which the probability of occurrence is maximized (solid vertical line, $\hat{\theta}$). Dashed vertical lines indicate the tolerance range ($\hat{\theta} - \hat{\sigma}$ to $\hat{\theta} + \hat{\sigma}$). The tolerance value is usually indicated by the estimated standard deviation ($\hat{\sigma}$).

Measures of spread in the function describe the species' tolerance. Typically, the standard deviation is used as the measure of tolerance (Jongman and ter Braak 1995), but the variance, intraquartile range, or other measures of spread could also be used. Standard deviation and variance are probably most convenient because standard approaches allow them to be derived easily from Gaussian functions (e.g., Equation 1).

Under some circumstances, Gaussian parameters can be derived relatively easily using simple averaging techniques (Charles 1985). Using this approach, species optima are calculated by averaging values of the environmental variable at sites where a species occurs (Salden 1978) or, alternatively, these environmental values can be averaged after weighting by species' abundances. However, if samples are not evenly distributed along the environmental gradient in which a species is found, if species' ranges are not limited to the edges of the environmental gradient, or if species are rare, weighted averaging may provide inaccurate estimates of optima and tolerance values (ter Braak and Looman 1986). Therefore, Gaussian logistic regression has been promoted for estimating optima and tolerance values for species exhibiting bell-shaped responses along environmental gradients (ter Braak and Looman 1986).

Gaussian logistic regression fits polynomial curves using link functions (transformations of the dependent variable) that prevent negative values in the function while maintaining the variance homogeneity assumptions of linear regression. The quadratic function that is used to describe species' responses is

$$\ln\left(\frac{p(x)}{1-p(x)}\right)=b_0+b_1x+b_2x^2=\frac{a-0.5(x-\theta)^2}{\sigma^2} \quad , \quad (1)$$

where θ is the species optimum, and σ is the species tolerance (ter Braak and Looman 1986).

The problem with this approach is that Gaussian models lack a strong theoretical foundation and may be inadequate for describing the response of a species to environmental conditions. Analyses of terrestrial plant populations suggest that Gaussian models often do an inadequate job of describing realized niches along environmental gradients (Austin 1976, 1980, 1999, 2002; Oksanen and Minchin 2002). These authors have argued that there is little justification for the universal application of the Gaussian function to describe species responses along environmental gradients.

If Gaussian models inadequately describe the environmental affinity of individual species, parameters

derived from Gaussian models may cause problems when used in environmental inference models. Inference models are used to estimate environmental conditions by averaging individual optima for species that are present at a site. Optima are generally derived from a calibration dataset in which species and environmental data are both available. The inference models are then used to estimate environmental conditions from diatom species composition. The ability to infer environmental conditions from species data is useful in fossil reconstructions of historic environmental conditions or other situations where environmental data are not available. Similar models have also been applied to infer conditions when environmental data are less reliable than desired, such as one time measurements of nutrients that are temporally variable (Stevenson, in preparation). If inaccurate parameters are used in these models, inaccurate or imprecise estimates of environmental conditions may result.

This study attempts to evaluate the ability of Gaussian models to accurately depict diatom species responses along gradients of pH, total phosphorus (TP), and total nitrogen (TN) by comparing Gaussian curves with flexible generalized additive models (GAM), which more closely track species' responses along these environmental gradients. This approach has been used in the past to evaluate parametric

model assumptions for a small number of terrestrial plant species (Oksanen and Minchin 2002), but this is the first extensive assessment of the Gaussian approach for diatom species, which are still commonly modeled using the Gaussian approach.

Additionally, inference models using optima and tolerance values derived from Gaussian models are compared with inference models using optima and tolerance values derived from GAM species' responses. Because GAMs include smoothed predictor variables, they track species responses more effectively than parametric Gaussian models. Therefore, inferences using GAM derived optima and tolerance values are expected to perform better than environmental inference models using Gaussian optima and tolerance values.

Mechanistic differences in the way that species respond to environmental stressors and resources (Grime 1973, 1977) may affect the ability of the Gaussian approach to model species responses to pH, TN, and TP. Thus, the effectiveness of the Gaussian approach is expected to differ between the three environmental gradients, particularly between pH and the two nutrient variables. Austin (1980) suggests that nutrient responses are likely to show complex shapes because the processes operating are likely to differ along the gradient. At the low end,

exploitative competition is likely to dominate, while at the upper end, a different resource may become limiting. If other resources remain sufficiently high, species responses might increase exponentially along the gradient until nutrients become toxic. Apparent competition may also result in the decline of a species at the high end of the nutrient gradient if increased grazer abundance is supported (Holt 1977). Provided that the gradient is long enough, species responses to pH are likely the result of optimal enzyme functioning. Therefore, species were expected to respond in a Gaussian fashion to pH, but not necessarily to nutrients, where one clear pattern was not expected to describe the responses of all species.

METHODS

Data

Diatom and environmental data used in this analysis came from streams in the Mid-Atlantic Highlands region of the eastern United States (Figure 2). Data were collected a part of the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program between 1993 and 1996. In several cases, stream reaches were sampled more than once. When this occurred, data for the site was pooled and environmental measurements were averaged across time to alleviate problems with repeated measures. This yielded a final set of 576 sites.

Algae samples were originally processed and identified in our lab and the data are now publicly available through the USEPA. Details regarding sampling, processing, and identification of diatom samples have been published elsewhere (Pan et al. 1996, 1999). A total of 628 diatom taxa were observed in the dataset; however, only those taxa observed at 10 or more sites were included in the assessment of Gaussian response curves, reducing the number of taxa to 204. Diatom taxa were generally identified at the resolution of species or variety.

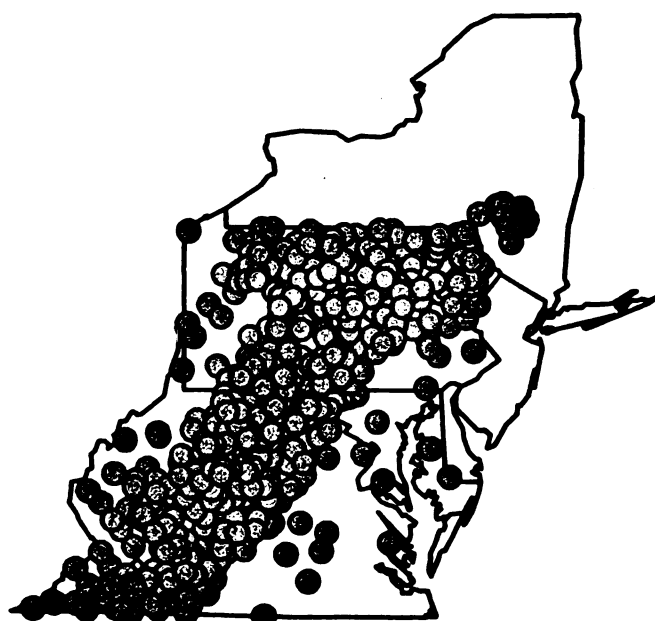


Figure 2. Stream sites sampled in the Mid-Atlantic Highlands region of the eastern United States as part of the United States Environmental Protection Agency's Environmental Monitoring and Assessment Program.

Total phosphorus and total nitrogen were measured using the persulfate oxidation and colorimetry method. Closed headspace measurements of pH were made within 72 hours of collection in air-tight syringes. Details regarding methods for chemical analysis is available through the USEPA (1987). Total phosphorus ranged from values below detection limit to 694 $\mu\text{g/L}$, with a mean of 25 $\mu\text{g/L}$. Total nitrogen ranged from 27 to 21730 $\mu\text{g/L}$ and averaged 916 $\mu\text{g/L}$. Stream pH had a minimum of 3.00 and a maximum of 8.82, averaging 7.58. Correlation between the chemistry values used in these analyses were statistically greater than zero in all cases, but none were high enough to cause concern when comparing the effectiveness of the Gaussian model between environmental gradients. The greatest correlation was 0.34 and existed between TN and TP.

Species Response Models and Evaluation

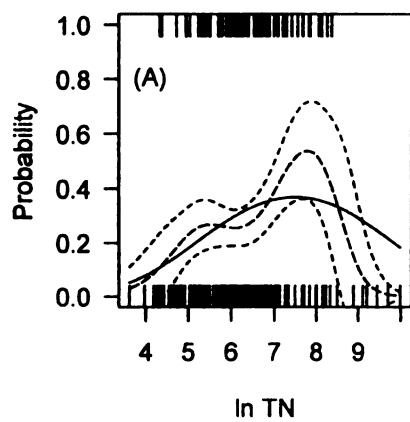
Binomial responses for each diatom taxon were modeled along environmental gradients of pH and natural-log-transformed TN and TP using generalized linear models (GLM) and GAM using a logit link function (Equation 1). Binomial responses were used in order to follow previous work described by ter Braak and Looman (1986) and Heegard (2002). Binomial responses are also easily interpreted as the probability that a species will be detected at a site

and are less sensitive to temporal fluctuations than absolute density or relative abundance. Gaussian model effectiveness was evaluated in three ways. First, if GLM predictions drifted outside of the GAM 95% confidence interval, the Gaussian models was considered a poor representation of the species' response (Figures 3a, 3b). Second, the Gaussian model was deemed inappropriate for deriving optima and tolerance levels if it was U-shaped (Figure 3b). In these cases, the method for deriving the mode of the Gaussian response places the environmental optimum at the least probable environmental condition to observe a species, rather than the most probable. Finally, if the predicted optimum fell outside of the range of observed environmental values (Figure 3c), the model was considered inappropriate because its parameters were outside of the observed universe of the dataset. If the Gaussian model for a species failed any of these three tests, the model was considered poor and if it passed all three tests (Figure 3d), it was considered effective at modeling the species response along a given environmental gradient.

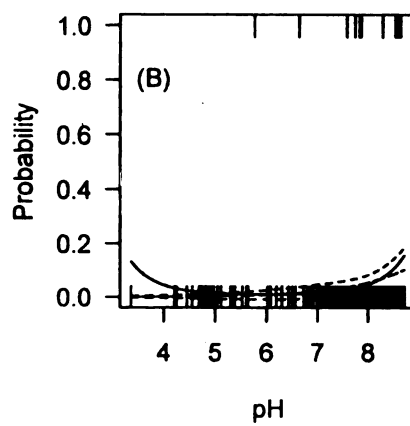
A bootstrapping method was applied to determine whether failure rates of the Gaussian approach for a given environmental gradient were unacceptably high. The sample population of 204 species for each environmental variable

Figure 3. Example diatom species response curves. Solid lines represent the Gaussian response curve estimated by GLM. Dashed lines represent GAM response curves and associated 95% confidence intervals. (a) Example of a GLM that failed because it was not contained within the 95% confidence interval of the GAM. (b) Example of a GLM that failed because it was U-shaped and because it was not contained within the 95% confidence interval of the GAM . (c) Example of a GLM that failed because the predicted optimum is outside of the observed range of the environmental variable. (d) Example of a GLM model that passed all evaluation criteria.

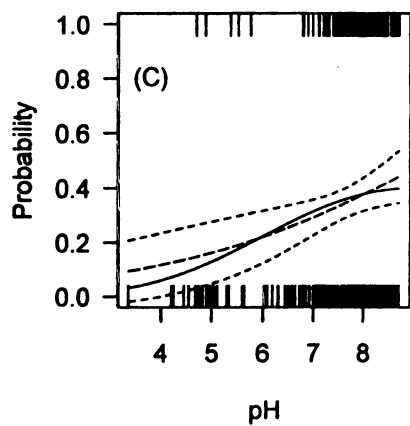
Nitzschia sociabilis



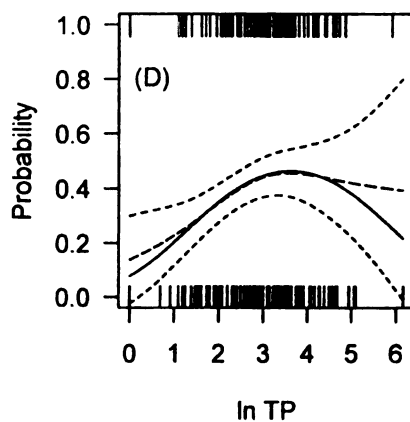
Cymbella reichardtii



Surirella angusta



Navicula cryptotenella



was re-sampled with replacement 10000 times to estimate sample variance of the failure rate. Rejection rate of the Gaussian method was determined at the 95% confidence level for arbitrarily chosen acceptable failure rates of 0.05 and 0.1. Therefore, if >95% of bootstrapped samples had failure rates below 0.1, the the Gaussian approach was deemed appropriate. If, however, <95% of bootstrapped samples had failure rates below 0.1, it could not be said with 95% confidence that the failure rate was below 0.1 and the Gaussian approach would not be considered appropriate.

To determine whether Gaussian failure rates differed significantly between the environmental gradients, paired differences in bootstrapped failure rates were calculated. If fewer than 5% of the paired differences were equal to 0 ± 0.05 , Gaussian failure rates between the two environmental parameters were considered significantly different.

Inference Models

Environmental conditions for each stream were inferred using the estimated optima and tolerance values of diatom taxa. Tolerance weighted estimates of environmental conditions were estimated as follows:

$$\hat{x}_i = \frac{\sum_{k=1}^m y_{ik} \hat{\theta}_k / \hat{y}_k}{\sum_{k=1}^m y_{ik} / \hat{y}_k} , \quad (2)$$

where $\hat{x}_i$ is the estimated environmental variable (pH, TN, or TP) for site i , y_{ik} is the proportional abundance of species k at site i , $\hat{\theta}_k$ is the mode of the GLM or GAM species response curve (i.e., the estimated optimum) for species k of m total species, and $\hat{y}_k$ is the estimated range of species k . The range is a measure of environmental amplitude and represents the span of conditions in which a species is expected to occur. It is calculated here following Heegaard (2002) as the difference between the upper and lower central borders of the species response curve (Figure 4). For binomial models using a logarithmic link function, the response value at the optimum, $E(g|x=\theta)$, is the mode, c , and the response value at the central borders is $E(g|x=\hat{\theta} \pm \sigma) = c \times e^{-0.5}$. For a normal Gaussian curve, the difference between the upper and lower central border is equivalent to $2 \times \hat{\sigma}_k$, where $\hat{\sigma}_k$ is the estimated standard deviation for species k .

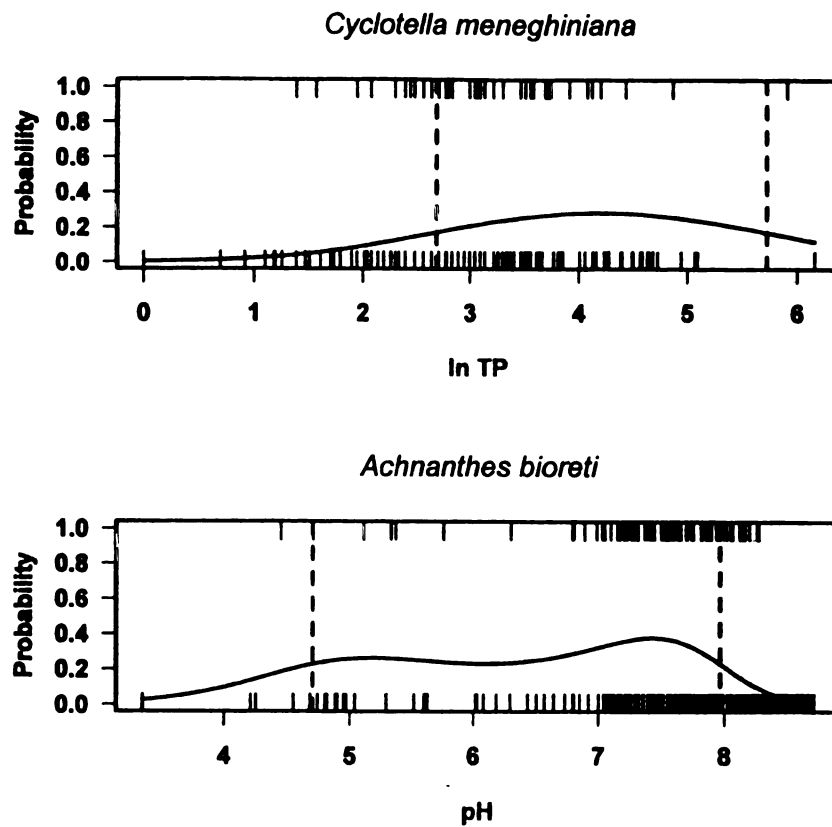


Figure 4. Examples of GLM (above) and GAM (below) response curves with estimated central borders (vertical dashed lines) from which species' tolerance ranges were calculated.

Inference models contain nested averages, biasing predictions toward the median of the observed range of environmental variables. This “shrinkage” of the environmental gradient is generally corrected using one of two deshrinking methods, classical deshrinking or inverse deshrinking. In both methods, deshrinking parameters, a constant (a) and a slope (b), are calculated by regressing observed and initial model predictions. In classical deshrinking, the initial estimates are regressed against the observed values to derive a and b (Equation 3a), while inverse deshrinking parameters are derived by regressing observed values on initial estimates (Equation 4a). Deshrinking corrected environmental estimates are calculated as follows, using these parameters:

$$initial \hat{x}_i = a + bx_i + \epsilon_i, \quad (3a)$$

$$final \hat{x}_i = (initial x_i - a) / b \quad (3b)$$

for classical deshrinking, and

$$\hat{x}_i = a + b \times initial \hat{x}_i + \epsilon_i, \quad (4a)$$

$$final \hat{x}_i = a + b \times initial x_i \quad (4b)$$

for inverse deshrinking.

The classical method deshrinks more than the inverse method, making it more effective at inferring values at the extreme ends of the environmental gradient (Birks et al. 1990). Due to the potential negative effects of

acidification and eutrophication on ecosystems, accurate prediction at low pH and high nutrient values are of particular interest. Therefore, classical desinking was used in this study.

Root mean squared error (RMSE) and correlation between observed and inferred values were used to assess fit and to compare models. The simpler approach assuming Gaussian species responses was considered better, unless inference models using GAM derived parameters showed more than a 5% increase in correlation and greater than 10% reduction in RMSE.

RESULTS

Evaluation of the Gaussian Model

At an acceptable failure rate of 5%, the Gaussian model inadequately described species responses along all three environmental gradients. Thirteen of the 204 models failed to adequately describe a species' response along the TP gradient (failure rate = 0.064) and had a bootstrapped failure probability of 0.7545. Total nitrogen models had a failure rate of 0.069, with a bootstrapped failure probability of 0.8386. The Gaussian model was also inadequate for pH, having a failure rate of 0.1863 and a bootstrapped failure probability of 1.0.

When the acceptable failure rate was increased to 10%, the Gaussian model passed for both TP and TN, which had bootstrapped probabilities of failure equal to 0.0226 and 0.0457, respectively. The Gaussian model failed for pH, even with the more liberal failure criteria of 10% ($\text{Pr}[\text{failure}]_{\text{boot}} = 1.0$).

Comparison of TP, TN, and pH Failure Rates

Failure rates for the two resource gradients, TP and TN, were not significantly different. Ninety-six percent of bootstrapped differences in failure rates were 0 ± 0.05 . When TP and TN failure rates were compared with pH,

however, each was significantly different. Ninety-nine percent of bootstrapped differences in failure rate between TP and pH were not equal to zero (± 0.05), while 98% of bootstrapped differences in failure rate between TN and pH were not equal to zero (± 0.05).

Performance of Inference Models

Models using the standard parameters derived from Gaussian GLM (Figures 5, 7, and 9) performed well when compared to inference using GAM derived parameters (Figures 6, 8, and 10). Inference using GAM parameters for TP and TN showed slight improvement over models using GLM parameters (Table 1). The magnitude of increase, however, was not significant enough to justify using the GAM parameters. The pH inference model actually performed worse using GAM parameters than when using GLM parameters, showing roughly a 3% decline in correlation and a 17% increase in prediction error.

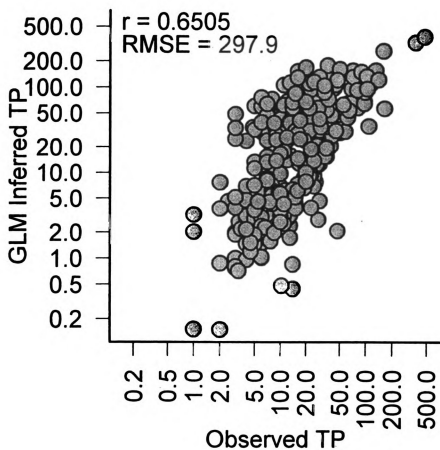


Figure 5. Observed vs. diatom-inferred total phosphorus in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.

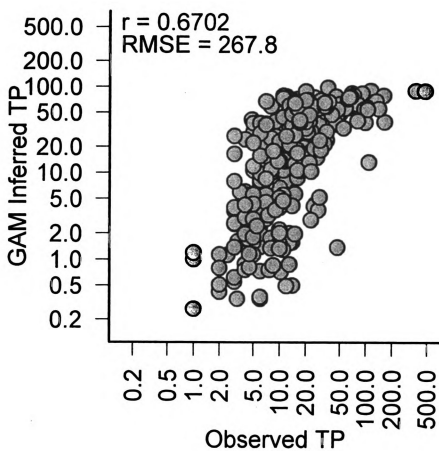


Figure 6. Observed vs. diatom-inferred total phosphorus in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.

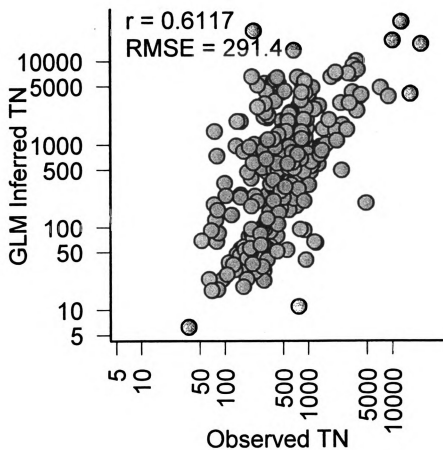


Figure 7. Observed vs. diatom-inferred total nitrogen in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.

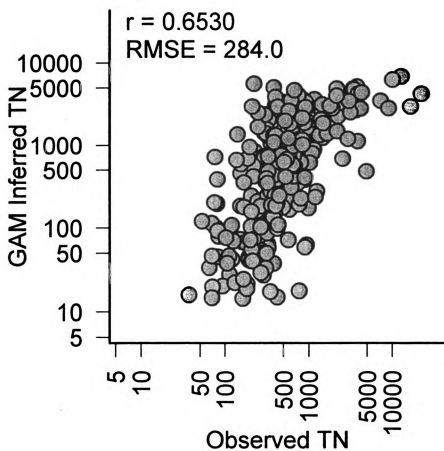


Figure 8. Observed vs. diatom-inferred total nitrogen in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.

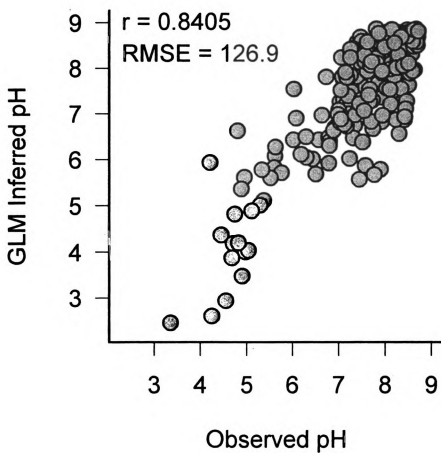


Figure 9. Observed vs. diatom-inferred pH in Mid-Atlantic Highlands streams. Generalized linear modeling (GLM) was used to derive species' optima and tolerance values used in the transfer function.

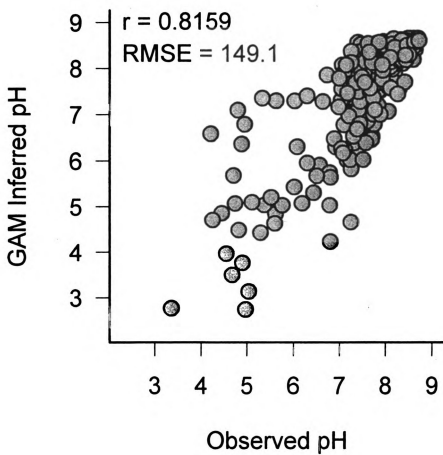


Figure 10. Observed vs. diatom-inferred pH in Mid-Atlantic Highlands streams. Generalized additive modeling (GAM) was used to derive species' optima and tolerance values used in the transfer function.

Table 1. Comparison of inference models using GLM and GAM derived parameters. A 5% increase in correlation and 10% decrease in RMSE was considered a significant improvement by using the GAM inference model.

	GLM		GAM		Percent Change	
	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE
TP	0.6505	297.9	0.6702	267.8	3.03%	-10.10%
TN	0.6117	291.4	0.6530	284.0	6.75%	-2.54%
pH	0.8405	126.9	0.8159	149.1	-2.93%	17.49%

DISCUSSION

The most common models used to infer environmental conditions with species autecological data make the assumption that species respond to environmental gradients in a Gaussian, bell-shaped, function. Although researchers have shown that the universal application of this model is inappropriate for many terrestrial plant species, it continues to be applied in paleolimnology and environmental assessment. The research described herein has attempted to 1) determine whether the Gaussian model is appropriate for benthic diatom species; 2) evaluate differences in species response functions along resource and stressor gradients, which mechanistically might be expected to differ; and 3) determine whether environmental inference models using parameters derived from smoothed response curves (GAM) perform significantly better than inference models that use parameters derived from Gaussian functions.

Counter to expectations, the results suggest that the standard generalized linear modeling approach may not work well for describing species' responses to pH, but that the Gaussian approach may work sufficiently well for TP and TN. When examining individual species' responses to the pH gradient, the Gaussian approach did a poor job of describing species' responses; Gaussian models failed for

nearly 20% of the diatom species examined. For nutrients, the Gaussian model performed better, failing for just over 6% of the diatom species examined.

Benthic diatom species appear to respond in a symmetric, Gaussian fashion along environmental gradients more often than other taxonomic groups that have been examined. In a study of terrestrial plant species, Oksanen and Minchin (2002) found that 45% of the species examined exhibited symmetric, bell-shaped responses along an altitude gradient. Their modeling approach used beta, rather than standard quadratic functions, but they suggest that symmetrical beta functions are likely to be modeled adequately by Gaussian models. Similarly, Minchin (1989) found that only 45 of 100 plant species' responses exhibited symmetric, unimodal responses along an altitude gradient. In a study of wetland and aquatic plant species, Bio et al. (1998) used stepwise multiple logistic regression to model species responses to several environmental variables simultaneously. They found that smoothed predictors (i.e., GAM) were included in 77% of 156 models; the Gaussian response curve fitted, on average, only 18% of responses. In another study of vascular plants, only 20% of species exhibited Gaussian responses along a pH gradient (Ejrnæs 2000). So, even along the pH gradient, which had the highest failure rate for diatom

species (19%), Gaussian models appear to be much more effective at describing species' responses for diatoms than for vascular plants.

Inference models using Gaussian response parameters also worked effectively for diatoms. Improvement in inference models for TP and TN using GAM derived optima and tolerance values was probably not large enough to justify the use of GAM over more traditional methods which are more easily computed. Interestingly, despite the poor performance of the Gaussian model for describing individual species' responses to pH, the inference model for pH using GAM derived parameters performed substantially worse than the model using Gaussian optima and tolerance values. The reason for this is not clear, but differences between observed and inferred values increased at the lower end of the gradient when GAM optima and tolerance values were used (Figures 10 and 11). These results suggest that standard methods for constructing inference models can be used effectively, despite potential flaws in the derivation of individual optima and tolerance values.

Failure rates between the pH stressor gradient and the nutrient resource gradients were significantly different. The Gaussian model was more effective at describing species' responses to resources than to the stressor. Mechanistically, such a difference in species responses to

stressors and to resources may be expected (Grime 1973, 1977). Stressors put physiological constraints on species, such as the ability of enzymes to function normally. Resources on the other hand should not have this effect until they reach toxic concentrations. At the low end of resources levels, competition is more likely to influence a species' ability to persist. Competition may lead to displacement of a species' optimum or even bimodal responses (Austin and Smith 1989). Therefore, diatom responses to nutrients and pH were expected to differ.

Austin, one of the most vocal opponents to the universal application of the Gaussian response curve, has suggested that there are three types of environmental gradients: indirect, direct, and resource gradients (Austin 1980) and that species responses along these types of gradients will differ. Indirect gradients do not have a direct effect on species performance. Indirect variables may correlate well with species performance in some areas, but not in others due to the presence of important covariates. Without accounting for these covariates, models constructed along indirect gradients are likely to be specific to one region. On the other hand, direct gradients such as pH have a direct physiological influence and may transfer well to other regions. Finally, resource gradients such as nutrients used by plants are either

sufficient or deficient, and the ability of a species to utilize low levels of a resource are unlikely to be correlated with the effects of oversupply. So, species' response are likely to be asymmetrical and the characteristics of the response curve will be sensitive to other species present in the community. Austin's theory regarding resource gradients suggests that Gaussian patterns might be expected along stressor gradients and non-Gaussian patterns might be expected for resource gradients. In this study, diatom species responded opposite to these expectations. Symmetrical, unimodal responses to TP and TN were observed, but pH optima were often predicted outside the range of observed values.

Diatom species' responses to pH may have differed from expectations due to indirect effects that exist in aquatic environments but are lacking from Austin's environmental gradients. Terrestrial plants are affected directly by soil pH, whereas aquatic producers may also be strongly affected indirectly by pH. Grazers are known to influence the abundance of algal species (Steinman 1996), and are also affected physiologically by pH (Baker and Christensen 1990). Furthermore, pH can influence resource availability (Fairchild and Sherman 1990). Therefore, non-Gaussian responses to pH may have been driven by indirect effects that are not expected for pH in terrestrial environments.

Despite their inadequate description of many species' responses to pH and several species' responses to TP and TN, the Gaussian function described diatom species' responses substantially better than it has for other taxonomic groups. Furthermore, environmental inference models using Gaussian species' response parameters performed nearly as well or better than inference models using GAM derived parameters. The ability of seemingly over simplistic inference models to effectively predict environmental conditions has been noted in the past (Birks 1998). Here it seems that more accurate descriptions of species' responses can even lead to less effective inference models. At this time, there appears to be little reason to abandon Gaussian approaches for modeling diatom species' responses, even though the universal application of this model lacks a strong theoretical foundation.

REFERENCES

- Austin, M. P. 1976. On non-linear species response models in ordination. *Vegetatio* 33: 33-41.
- Austin, M. P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11-21.
- Austin, M. P. 1999. A silent clash of paradigms: some inconsistencies in community ecology. *Oikos* 86: 170-178.
- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101-118.
- Austin, M. P., and Smith, T. M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.
- Baker, J. P., and Christensen, S. W. 1990. Effects of acidification in biological communities in aquatic ecosystems. Pages 83-106 in Charles, D. F., ed. *Acid Deposition and Aquatic Ecosystems: Regional Case Studies*, Springer-Verlag, New York, NY.
- Battarbee, R. W., and Charles, D. F. 1987. The use of diatom assemblages in lake-sediments as a means of assessing the timing, trends, and causes of lake acidification. *Progress in Physical Geography* 11: 552-580.
- Bio, A. M. F., Alkemade, R., and Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *Journal of Vegetation Science* 9: 5-16.
- Birks, H. J. B. 1998. D. G. Frey and E. S. Deevey Review #1: Numerical tools in paleolimnology—Progress, potentialities, and problems. *Journal of Paleolimnology* 20: 307-332.
- Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., and ter Braak, C. J. F. 1990. Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London B* 327: 263-278.

- Charles, D. F. 1985. Relationships between surface sediment diatom assemblages and lakewater characteristics in Adirondack lakes. *Ecology* 66: 994-1011.
- Charles, D. F., and Smol, J. P. 1988. New methods for using diatoms and chrysophytes to infer past pH of low-alkalinity lakes. *Limnology and Oceanography* 33: 1451-1462.
- Dixit, S. S., Smol, J. P., Kingston, J. C., and Charles, D. F. 1992. Diatoms: powerful indicators of environmental change. *Environmental Science and Technology* 26: 22-33.
- Ejrnaes, R. 2000. Can we trust gradients extracted by detrended correspondence analysis. *Journal of Vegetation Science* 11: 565-572.
- Fairchild, G. W., and Sherman, J. W. 1990. Effects of liming on nutrient limitation of epilithic algae in an acidic lake. *Water, Air and Soil Pollution* 52: 133-147.
- Grime, J. P. 1973. Control of species density in herbaceous vegetation. *Journal of Environmental Management* 1:151-167.
- Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist* 111: 1169-1194.
- Heegard, E. 2002. The outer border and central border for species-environmental relationships estimated by non-parametric generalized additive models. *Ecological modeling* 157: 131-139.
- Holt, R. D. 1977. Predation, apparent competition and the structure of prey communities. *Theoretical Population Ecology* 12: 197-229.
- Jongman, R. G., and ter Braak, C. J. 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge, UK.
- Krebs, C. J. 1994. *Ecology: the experimental analysis of the distribution and abundance*, 4th edition. Harper and Row, New York.
- Minchin, P. 1989. Montane vegetation of the Mt. Field Massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio* 83: 97-110.

- Oksanen, J. and Minchin, P. R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157: 119-129.
- Pan, Y., Stevenson, R. J., Hill, B. H., Herlihy, A. T., and Collins, G. B. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. *Journal of the North American Benthological Society* 15: 481-495.
- Pan, Y., Stevenson, R. J., Hill, B. H., Kaufmann, P. R., and Herlihy, A. T. 1999. Spatial patterns and ecological determinants of benthic algal assemblages in Mid-Atlantic Streams, USA. *Journal of Phycology* 35: 460-468.
- Ricklefs, R. E. 1997. *The Economy of Nature*, 4th edition. W. H. Freeman and Company, New York.
- Salden, N. 1978. Beiträge zur Ökologie der Diatomeen (Bacillariophyceae) des Süßwassers. *Decheniana, Beiheft* 22: 1-238.
- Steinman, A. D. 1996. Effects of grazers on freshwater benthic algae. Pages 341-373 in Stevenson, R. J., Bothwell, M. L., and Lowe, R. L., eds. *Algal Ecology: Freshwater Benthic Ecosystems*, Academic Press, San Diego, CA.
- ter Braak, C. J. F., and Looman, C. W. M. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65: 3-11.
- United States Environmental Protection Agency (USEPA). 1987. *Handbook of Methods for Acid Deposition Studies, Laboratory Analysis for Surface Water Chemistry*. EPA 600/4-87/026. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

CHAPTER THREE

PREDICTING DIATOM ASSEMBLAGES IN MINIMALLY DISTURBED STREAMS USING A NEW HYBRID MODELLING APPROACH

INTRODUCTION

Determining whether the ecological integrity of a stream reach has been affected by human activities requires establishing reference conditions as a "control". Because historical data are rarely available, the traditional approach is to compare the ecological conditions of a stream reach to a site upstream of the suspected cause of impairment (Lowe and Pan 1996, Hansmann and Phinney 1973). The weakness of this approach is the lack of independence between samples (Hurlbert 1984), lack of replication to account for uncertainty, potential lack of adequate reference sites upstream, and difficulty addressing problems associated with non-pointsource pollutants.

In an attempt to address several of these problems, many agencies in the United States are evaluating or have adopted a regionalization approach (Barbour et al. 1996, DeShon 1995, Yoder and Rankin 1995, Ohio EPA 1987). In this approach, streams within a spatially contiguous region are expected to be similar in the absence of human disturbance. Reference sites with minimal human disturbance are then chosen within each region. Test sites (sites being evaluated for impairment) are then compared to reference sites within the region to determine whether conditions are significantly different than expected.

While improving problems associated with replication, sample independence, and non-pointsource pollutants, the regionalization approach can have weaknesses associated with assumed site homogeneity within the regionalization classes. Hydrogeologically heterogeneous regions like those found in the western United States can lead to substantial variation in biological conditions within a region, even with minimal human disturbance.

An approach applied in the United Kingdom (Clarke et al. 2003) and Australia (Norris and Norris 1995) attempts to overcome this problem by classifying streams according to their biology, and does not assume spatial continuity within classes. In RIVPACS-type models (River InVertebrate Prediction And Classification System, Moss et al. 1987), reference sites are chosen a priori and biological assemblages are sampled in these reference sites. Multivariate statistical clustering techniques are then used to classify streams according to the biological data. Inference models are then developed to predict site membership using hydrologic, geologic, geographic, and climate variables that are relatively independent of human activities at the regional or subcontinental scale. The observed biological community at a test site is then compared to the inferred biological community predicted to exist at the site.

Linear discriminant analysis (LDA; a.k.a. multiple discriminant analysis, MDA, discriminant functions analysis, DFA, or discriminant analysis, DA) has been used to build RIVPACS-type models to predict expected conditions at a site using environmental predictor variables. Linear discriminant analysis attempts to discriminate between predefined, discrete categories of sampling units, maximizing the among-to-within group variation using linear combinations of variables (McCune and Grace 2002).

The application of LDA in ecological assessment poses several potential problems related to the assumptions of this technique. Linear discriminant analysis assumes within-group variance homogeneity, multivariate normality, and linear relationships between variables (McCune and Grace 2002). Furthermore, LDA lacks an effective way to handle missing data. Ecological and environmental data often fail to meet these assumptions. Chemical concentrations and species abundances are often left-censored (i.e., below detection limits), many ecological patterns are known to be non-linear, and missing data are common in the large datasets used to develop predictive models. Thus, the assumptions of LDA, combined with the attributes of ecological data, may result in imprecise or inaccurate predictive models (Williams 1983).

One alternative to LDA when predicting group membership

1
T.
at
th.

for new observations is recursive partitioning (a.k.a. classification and regression trees or CART; Breiman et al. 1984). Recursive partitioning alleviates many of the problems associated with LDA because it allows for many data types, including continuous and categorical predictor variables; it is insensitive to non-normal variance distributions of the variables; and it models relationships in a non-linear and hierarchical fashion. While LDA seeks to maximize among-to-within group variation using linear combinations of variables, CART seeks to split data into increasingly homogeneous groups by repeatedly splitting data using a single value of one predictor variable that minimizes impurity within each of the resulting groups. For classification trees (i.e., trees with categorical responses), impurity in the resulting groups may be evaluated using one of several possible indices, the most common being the information index, the Gini index, and the twoing index (Breiman et al. 1984). Each node containing a set of observations is recursively split into binary child nodes until a stopping rule is reached (e.g., a predefined minimum improvement in the variance explained or a minimum number of observations within each of the terminal nodes). Thus, CART finds changepoints rather than linear patterns and approaches pattern recognition in a hierarchical rather than an additive manner. Despite its effective handling of

many problematic forms of environmental data, CART has its limitations. Recursive partitioning is inefficient at modeling linear relationships when they do exist.

Approaches combining LDA and CART may help exploit the advantages of each technique, while minimizing the effects of their weaknesses. Such approaches have been demonstrated for characterizing species' traits related to invasion success and in assessing the risk of invasion by these species (Rejmanek and Richardson 1996, Reichard and Hamilton 1997, Kolar and Lodge 2002). These approaches, however, have synthesized LDA and CART model predictions in an informal, qualitative way. For example, a species might be considered an invasion threat if both LDA and CART predict the species to be a threat, regardless of the probabilities associated with those classifications. Perhaps more importantly, these approaches provide little guidance regarding the interpretation of results when the two model predictions disagree.

In this paper, an inference model for predicting diatom species composition under minimally disturbed conditions in streams throughout the western United States is presented. The inference model uses a new method that formally integrates the predictions of LDA and CART using Bayes' theorem. Misclassification rates and the ratio of correct-to-incorrect predictions weighted by classification

probabilities are compared for LDA, CART, and the new hybrid approach.

METHODS

Data

The data used in these analyses were collected from reference sites throughout the western United States (Figure 11). We sampled reference sites that were generally minimally disturbed by human activities and often nearly pristine. Including streams that were representative of most wadable stream types within a region sometimes required sampling streams that were the best available. Site selection prior to sampling was based on best professional judgment of local aquatic scientists and by landscape features. Posterior elimination of sites from the reference dataset occurred when field crews noted significant anthropogenic degradation of a site, relative to similar stream types.

Physical, chemical, and biological data were collected at each site. To minimize the confounding of spatial and temporal patterns in the data, sites were not sampled in a spatially systematic fashion and five field crews were sent to separate locations throughout the western United States. Full descriptions of collection protocols are available through the Western Center for Monitoring and Assessment of Freshwater Ecosystems (<http://www.cnr.usu.edu/wmc/>) and are forthcoming in publications elsewhere. For brevity, only a

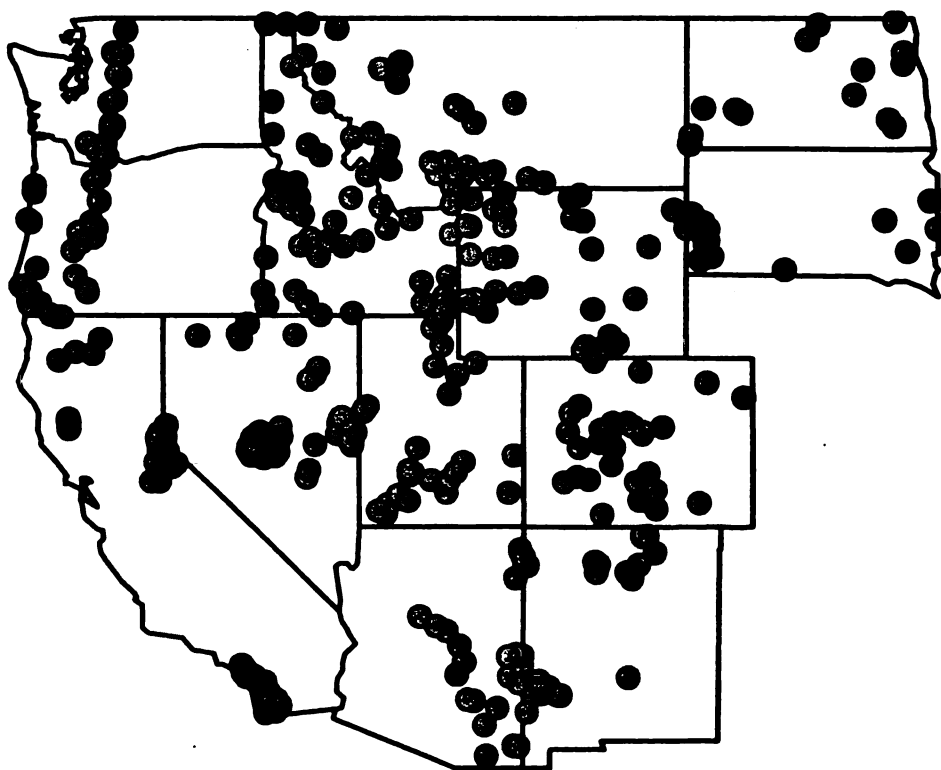


Figure 11. Location of stream reaches sampled throughout the western United States.

subset of the protocols are described here, with emphasis on periphyton collection. Sampling reaches were composed of four fast-water habitats (i.e., riffle, run, or rapid) at a study site. Within each of the 4 fast-water habitats, 4 rocks were collected, and periphyton was scraped from a delimited area of roughly 15 cm². On rare occasions, rocks were not available and other substrata were sampled, prioritized from bedrock/boulder, to sand and silt. Periphyton from all substrata were composited, subsampled, and preserved with formalin.

In the lab, I homogenized periphyton samples and digested subsamples in hot nitric acid to clear diatom frustules of organic matter. A subsample of cleaned diatom frustules were then permanently mounted on slides in Naphrax(R) and I identified individual diatom valves within transects to the lowest reasonable taxonomic level using 1000x total magnification. Generally diatoms were identified to the species or variety level. Computationally, a random subsample of 500 valves for each site was selected to ensure an equal sample size across sites.

We sampled a total of 408 sites that were included in this analysis during the summers of 2001 and 2002. Predictor variables were chosen to be relatively independent of human influence and included both field-

based and GIS-derived data. These variables included a hydrologic retention index determined from the difference in time between the leading and trailing edges of a fluorescent dye release along 50 m to 100 m stretch of the sampling reach (7.87 ± 29.53 , mean $\pm$ standard deviation), percent canopy cover ($45\% \pm 33\%$), percent stream slope ($4\% \pm 4\%$), stream velocity (0.36 ± 0.29 m/s), Julian date (from day 135 to day 254), percent of substrata >128 mm in diameter ($23\% \pm 19\%$), average high values of bulk soil density within the basin (1.44 ± 0.14 g/cm³), latitude (40.8002 ± 4.3348 degrees north), longitude (-113.0856 ± 6.3249 degrees west), elevation (1638 ± 786 m), watershed area (777 ± 3574 km²), average high values of organic matter content in basin soils ($2\% \pm 1\%$ w/w), average high values of basin soil permeability (12.4 ± 9.0 cm/h), mean monthly precipitation (917 ± 575 mm), average relative humidity (56.3 ± 8.5), average high values of depth to bedrock (46.3 ± 9.9 m), mean annual maximum monthly temperature ($12.6 \pm 4.4^{\circ}\text{C}$), mean annual monthly temperature ($5.7 \pm 4.1^{\circ}\text{C}$), mean annual minimum monthly temperature ($-1.2 \pm 4.0^{\circ}\text{C}$), average number of days with measurable precipitation (97 ± 37 days), and the ratio of minimum mean monthly flow to maximum mean monthly flow interpolated from USGS gaging stations as an index of hydrologic stability (0.03 ± 0.04).

8

/

da

le

Clustering Sites by Taxonomic Composition

Various clustering techniques have been applied by researchers using RIVPACS-type models. Some have found success using two-way indicator species analysis (TWINSpan, Hill 1979; Clarke et al. 2003, Moss et al. 1987), while others have preferred agglomerative hierarchical clustering methods (Ostermiller and Hawkins 2004, Hawkins et al. 2000). An artificial neural network technique called a self-organizing map (SOM) has been used by others to classify stream sites using diatom assemblages (Gevrey et al. 2004). While any of these clustering methods will work with the predictive modeling technique described below, SOM was used in this study to classify sites using diatom genus composition. Preliminary analyses suggested that clusters were more homogeneous when taxonomic resolution was reduced to the genus level, so sites were classified using diatom genus composition. Self-organizing maps were chosen because of their previous application to diatom assemblages (Gevrey et al. 2004) and the proven ability of artificial neural networks to handle complex non-linear data found in ecological data (Lek et al. 1999, Recknagel et al. 1997). Sites were clustered by genera using the *som* package in R (R Development Core Team 2005). Multidimensional genus data was mapped onto a 5x4 hexagonal grid following natural log plus one transformation of genus counts. Each SOM cell

was treated as a discrete class.

Predicting Class Membership

Initially, two predictive models, one using LDA and the other using CART, were developed to predict site membership using environmental variables. Linear discriminant analysis was performed using the *lda* function in the *MASS* library for R. In the LDA model, predictor variables were centered and scaled by their standard deviation. Sites with missing variables were dropped when building the LDA model. Sites were probabilistically assigned to each SOM class. In cases where sites were dropped due to missing variables, prior probabilities of class membership were used. A prior class probability was determined from the proportion of all 408 sites that belonged to the class. Probabilistic classifications for the LDA model were calculated using a leave-one-out jack-knifing procedure to prevent overestimation of probabilities that may result from resubstitution of the training set.

A classification tree was developed using the same predictors that were used in the LDA model. However, this is not a requirement of the hybrid method described below and was only done here for the purpose of comparing LDA, CART, and the hybrid models. In application, the best predictors should be selected for each model. It is likely

R

R

R

cl

pr

Ve

that these predictors will differ between LDA and CART because the best variables for LDA should be linear, while the best variables for CART should result in strong breakpoints. The *rpart* library of *R* was used to grow and select the classification tree. Split impurity was evaluated using the Gini index. *V*-fold cross-validation (*V*=10) was used to select the appropriate tree size. In this method, the data are split into *V* sub-groups of approximately equal size. Each group is then removed, one group at a time, and the remaining data is used to construct the model. The model is then used to predict class membership for sites in the removed sub-group. The relative error is calculated for every possible tree size for each of the *V* models. The relative error for each tree size is then summed across all *V* models and the tree size with the smallest total relative error is chosen (Breiman et al. 1984).

Bayes' rule can be used to combine the LDA and CART model predictions. For each new observation $\tilde{y}$ classified by the LDA and CART models, there exists a vector of probabilities that the observation belongs to each of the *k* possible classes. The maximum likelihood approach classifies $\tilde{y}$ in the group that has the greatest probability of membership. The classification probability vectors resulting from LDA and CART are easily combined

using Bayes' rule, which provides a framework to integrate LDA and CART predictions in a formal way. Bayes' rule states the probability of a new observation, $\tilde{y}$, belonging to class k , conditional on the data at hand, y , is proportional to the likelihood of the data, $Pr(y|\tilde{y}_k)$, times the prior probability, $Pr(\tilde{y}_k)$. This value is then divided by a normalization constant equal to the sum of all k products of the likelihood and the prior, bringing the sum of all $Pr(\tilde{y}_k|y)$ to one (Equation 5).

$$Pr(\tilde{y}_k|y) = \frac{Pr(y|\tilde{y}_k) \times Pr(\tilde{y}_k)}{\sum_{k=1}^K Pr(y|\tilde{y}_k) \times Pr(\tilde{y}_k)} \quad (5)$$

Thus, multiplying the probability vectors output by LDA and CART leads to posterior probabilities that incorporate the predictions of both models. The sum of all probable outcomes for some event must equal one. The denominator of Equation 1 accomplishes this for the updated probabilities, dividing each of them by the sum of all products for that observation.

In this approach, either the LDA or the CART predictions represent the prior, while predictions of the other model represent the likelihood and the two are interchangeable,

$$Pr(\tilde{y}_k|y) = \frac{Pr(y|\tilde{y}_k) \times Pr(\tilde{y}_k)}{\sum_{k=1}^k Pr(y|\tilde{y}_k) \times Pr(\tilde{y}_k)} = \frac{Pr(\tilde{y}_k) \times Pr(y|\tilde{y}_k)}{\sum_{k=1}^k Pr(\tilde{y}_k) \times Pr(y|\tilde{y}_k)} \quad (6)$$

The two sets of prediction probabilities are different, however, in the way they were derived and, in practice, they probably differ in the variables used to create them.

Maximum likelihood estimates of the hybrid, LDA, and CART modeling approaches were then compared using the percent of misclassified observations and a confidence weighted ratio which takes into account the predicted probability of membership for correctly and incorrectly classified observations. A standard method for evaluating these types of predictive classification models is to resubstitute the observations used to create the model and to quantify the percent of misclassified observations (Moss, 2000). One problem with this evaluation approach is that it does not take into account the amount of confidence the model has placed in each classification. Conceivably, an observation in a model with 20 categories could be placed into a category having a probability 0.051, if all other categories have probabilities ≤ 0.05 . It is very likely that the site is misclassified. Confidence in the prediction is quite low, but this is not taken into account using misclassification rates. Therefore, model predictions were also compared using a confidence-weighted

ratio (*CWR*) of correct to incorrect classifications

$$CWR = \frac{\sum_{i=1}^n \max(Pr_i|correct)}{\sum_{i=1}^n \max(Pr_i|incorrect)} , \quad (7)$$

where $\max(Pr_i|correct)$ is the value of the maximum probability for observation i , given that the classification was correct. Likewise, $\max(Pr_i|incorrect)$ is the value of the maximum probability for observation i , given that the classification was incorrect.

RESULTS

Clustering of Sites

The self-organizing map resulted in 19 of 20 possible classes. Most classes contained 8 or more observations, but several contained fewer due to unique diatom assemblages (Table 2). Smaller grid dimensions were tried, but this led to excessive lumping of larger classes and did little to incorporate smaller classes into larger ones. One cell in the 5x4 map (C21) did not contain any sites. A small number of genera were common to at least one site within every class. These genera included *Achnantheidium*, *Amphora*, *Cocconeis*, *Diatoma*, *Fragilaria*, *Navicula*, *Nitzschia*, *Planothidium*, and *Staurosirella*. More commonly, genera were unique to only a small number of classes (Table 3).

LDA Model Summary

The first linear discriminant axis accounted for 49.2% of the variation, while axes 2 and 3 accounted for just over 10% each. Temperature-related variables were the most important for discriminating among diatom genus classes using linear combinations of variables. Mean annual monthly temperature was positively associated with the first discriminant axis, while minimum and maximum annual

Table 2. Self-organizing map classification summary, including the number of sites place in each class (N) and the associated mean squared error of each class (MSE).

Class	N	MSE
C00	82	7.04
C01	4	6.2
C02	27	6.25
C03	55	6.29
C10	8	6.81
C11	2	7.08
C12	2	7.74
C13	28	6.22
C20	13	7.25
C22	2	5.09
C23	30	6.8
C30	14	6.88
C31	1	7.43
C32	3	6.15
C33	19	6.74
C40	28	7.1
C41	35	6.92
C42	2	5.9

Table 3. Proportion of sites within each class in which diatom genera were observed.

Genus	Proportion of Sites Per Class									
	C00	C01	C02	C03	C10	C11	C12	C13	C20	C22
<i>Achnanthidium</i>	0.8	1	1	0.98	1	1	1	1	1	1
<i>Achnanthes</i>	0.05	0.5	0.19	0.16			0.5	0.14	0.23	0.5
<i>Actinocyclus</i>	0.02									
<i>Adlafia</i>	0.1	0.25	0.22	0.56	0.13		0.5	0.54	0.15	
<i>Amphipleura</i>	0.2		0.07	0.04	0.13				0.15	0.5
<i>Amphora</i>	0.89	1	0.81	0.71	1	0.5	0.5	0.79	0.77	1
<i>Aneumastus</i>										
<i>Anomoeoneis</i>										
<i>Asterionella</i>			0.04	0.02				0.04		
<i>Aulacoseira</i>	0.16	0.25	0.15	0.11	0.13		0.5	0.18		
<i>Bacillaria</i>	0.06			0.02						
<i>Biremis</i>	0.06									
<i>Brachysira</i>				0.02						
<i>Caloneis</i>	0.4		0.33	0.18	0.5			0.04	0.31	
<i>Campylodiscus</i>										
<i>Cavinula</i>			0.04	0.02				0.07		
<i>Chamaepinnularia</i>	0.04		0.04	0.02						
<i>Cocconeis</i>	0.98	1	1	1	1	1	1	1	0.92	1
<i>Craticula</i>	0.32	0.5	0.33	0.38	0.5			0.32	0.23	0.5
<i>Ctenophora</i>	0.02									
<i>Cyclostephanos</i>	0.1									
<i>Cyclotella</i>	0.48		0.22	0.15	0.13	0.5		0.18	0.31	
<i>Cymatopleura</i>	0.01			0.02						0.5
<i>Cymbella</i>	0.24	0.75	0.22	0.31	0.5	1		0.39	0.62	1
<i>Denticula</i>	0.1				0.38				0.15	
<i>Desmogonium</i>	0.01									
<i>Diadesmis</i>		0.25	0.15	0.24				0.11		
<i>Diatomella</i>			0.04	0.05				0.04		
<i>Diatoma</i>	0.29	0.75	0.44	0.67	0.63	1	0.5	0.93	0.31	0.5
<i>Didymosphenia</i>	0.01			0.04						
<i>Diploneis</i>	0.26		0.15	0.13		0.5		0.11	0.31	
<i>Encyonema</i>	0.35	1	0.56	0.76	0.38	1		0.96	0.69	1
<i>Encyonopsis</i>	0.13	0.5	0.15	0.15	0.63			0.11	0.23	
<i>Entomoneis</i>										

Table 3 (con't)

Genus	Proportion of Sites Per Class								
	C23	C30	C31	C32	C33	C40	C41	C42	C43
<i>Achnanthidium</i>	1	1	1	1	1	1	1	1	1
<i>Achnanthes</i>	0.2	0.29			0.26	0.25	0.6		0.62
<i>Actinocyclus</i>									
<i>Adlafia</i>	0.37	0.14	1	0.67	0.42	0.14	0.14	1	0.36
<i>Amphipleura</i>	0.03	0.21			0.05	0.18	0.03		0.04
<i>Amphora</i>	0.6	1	1	1	0.53	0.82	0.31	1	0.49
<i>Aneumastus</i>		0.07							
<i>Anomoeoneis</i>							0.03		0.02
<i>Asterionella</i>					0.05	0.04	0.03		0.08
<i>Aulacoseira</i>	0.23				0.21		0.29		0.42
<i>Bacillaria</i>									
<i>Biremis</i>									
<i>Brachysira</i>	0.07	0.14			0.11	0.11	0.17		0.06
<i>Caloneis</i>	0.17	0.5			0.05	0.43	0.06		0.21
<i>Campylodiscus</i>		0.07							
<i>Cavinula</i>									0.08
<i>Chamaepinnularia</i>	0.07								0.04
<i>Cocconeis</i>	1	1	1	1	0.84	0.79	0.71	1	0.92
<i>Craticula</i>	0.13	0.14			0.32		0.03		0.06
<i>Ctenophora</i>						0.04			0.02
<i>Cyclostephanos</i>									
<i>Cyclotella</i>	0.07			0.33	0.05	0.11	0.09	0.5	0.06
<i>Cymatopleura</i>									
<i>Cymbella</i>	0.17	0.79	1	0.67	0.37	0.75	0.51	0.5	0.45
<i>Denticula</i>		0.14				0.21			0.02
<i>Desmogonium</i>									
<i>Diadesmis</i>	0.2				0.16	0.04			0.13
<i>Diatomella</i>	0.03								
<i>Diatoma</i>	0.8	0.57	1	0.67	0.84	0.5	0.49	1	0.83
<i>Didymosphenia</i>				0.33	0.05	0.04	0.03		0.08
<i>Diploneis</i>	0.07	0.14				0.25	0.03		
<i>Encyonema</i>	0.73	0.86	1	1	0.84	0.75	0.71	1	0.98
<i>Encyonopsis</i>	0.1	0.57	1		0.05	0.79	0.37		0.13
<i>Entomoneis</i>						0.04			

Table 3 (con't)

Genus	Proportion of Sites Per Class									
	C00	C01	C02	C03	C10	C11	C12	C13	C20	C22
<i>Eunotia</i>	0.13		0.07	0.15				0.21		
<i>Fallacia</i>	0.04	0.25		0.05	0.13				0.08	
<i>Fragilaria</i>	0.68	1	0.81	0.87	0.5	0.5	0.5	0.93	0.77	1
<i>Fragilariforma</i>				0.02						
<i>Frustulia</i>	0.11		0.15	0.13			0.5	0.25	0.08	0.5
<i>Geissleria</i>	0.29	0.5	0.74	0.75	0.5	0.5	1	0.61	0.54	0.5
<i>Gomphosphenia</i>	0.01		0.04	0.04			0.5	0.11		
<i>Gomphoneis</i>	0.16		0.22	0.27				0.29	0.15	
<i>Gomphonema</i>	0.93	0.75	1	0.98	1	1	1	1	0.85	1
<i>Gyrosigma</i>	0.04		0.04					0.04	0.15	
<i>Hannaea</i>	0.04	0.25		0.29				0.57	0.08	0.5
<i>Hantzschia</i>	0.04		0.15	0.11				0.07	0.08	
<i>Hippodonta</i>	0.17		0.04	0.04	0.25	0.5		0.04		
<i>Karayevia</i>	0.1		0.19	0.16	0.38		0.5	0.07	0.23	
<i>Kolbesia</i>	0.01		0.04	0.2			0.5	0.14		
<i>Lemnicola</i>	0.02		0.04		0.13			0.04		
<i>Luticola</i>	0.02		0.07	0.11				0.04		
<i>Martyana</i>	0.01									
<i>Mastogloia</i>	0.01									
<i>Mayamaea</i>	0.34	0.5	0.63	0.65	0.38	0.5		0.5	0.15	1
<i>Melosira</i>	0.22	0.25	0.15	0.15	0.25		0.5	0.18	0.15	
<i>Meridion</i>	0.17	0.25	0.19	0.4	0.25	1		0.5	0.15	
<i>Microcostatus</i>			0.04							
<i>Navicula</i>	1	1	1	0.98	1	1	1	0.96	1	1
<i>Neidium</i>			0.04	0.02	0.13			0.04		
<i>Nitzschia</i>	0.99	1	1	0.98	1	1	0.5	1	0.92	1
<i>Nupela</i>				0.2			0.5	0.04		
<i>Opephora</i>	0.01	0.25	0.11	0.04				0.04		
<i>Orthoseira</i>										
<i>Parlibellus</i>										
<i>Pinnularia</i>	0.07	0.25	0.15	0.18				0.18	0.08	
<i>Placoneis</i>	0.02			0.04				0.04		

Table 3 (con't)

Genus	Proportion of Sites Per Class								
	C23	C30	C31	C32	C33	C40	C41	C42	C43
<i>Eunotia</i>	0.27				0.16	0.04	0.4		0.32
<i>Fallacia</i>		0.29				0.14			
<i>Fragilaria</i>	0.93	0.5	1	0.67	0.95	0.64	0.77	0.5	0.94
<i>Fragilariforma</i>					0.05				
<i>Frustulia</i>	0.17	0.21				0.04	0.2		0.17
<i>Geissleria</i>	0.67	0.43	1	1	0.42	0.32	0.11		0.4
<i>Gomphosphenia</i>	0.1						0.09		0.13
<i>Gomphoneis</i>	0.4	0.07		0.33	0.37	0.07	0.06	0.5	0.19
<i>Gomphonema</i>	0.97	1		1	1	0.82	0.89	1	1
<i>Gyrosigma</i>		0.14				0.07			
<i>Hannaea</i>	0.63		1		0.68	0.14	0.51	0.5	0.87
<i>Hantzschia</i>	0.07					0.04			0.06
<i>Hippodonta</i>									
<i>Karayevia</i>	0.07		1		0.05	0.11		0.5	0.13
<i>Kolbesia</i>	0.5			0.33	0.37		0.17		0.28
<i>Lemnicola</i>									
<i>Luticola</i>	0.03				0.16	0.04	0.03		0.02
<i>Martyana</i>									
<i>Mastogloia</i>						0.07			
<i>Mayamaea</i>	0.5	0.21		0.67	0.53	0.04	0.09	1	0.17
<i>Melosira</i>	0.27	0.07			0.11		0.06		0.17
<i>Meridion</i>	0.4	0.36			0.42	0.14	0.26		0.55
<i>Microcostatus</i>									
<i>Navicula</i>	0.97	1	1	1	1	0.96	0.63	1	0.87
<i>Neidium</i>	0.03	0.07				0.04	0.03		
<i>Nitzschia</i>	0.97	1	1	1	0.89	0.93	0.6	1	0.81
<i>Nupela</i>	0.13			0.33	0.21		0.06		0.15
<i>Opephora</i>	0.07						0.03		0.06
<i>Orthoseira</i>									0.02
<i>Parlibellus</i>									0.02
<i>Pinnularia</i>	0.13	0.07			0.05	0.04	0.06		0.13
<i>Placoneis</i>		0.07			0.05				

Table 3 (con't)

Genus	Proportion of Sites Per Class									
	C00	C01	C02	C03	C10	C11	C12	C13	C20	C22
<i>Pseudostaurosira</i>	0.17		0.15	0.18	0.38			0.29	0.15	
<i>Punctastriata</i>										
<i>Reimeria</i>	0.5	1	0.81	0.91	0.88	1		0.89	0.46	1
<i>Rhoicosphenia</i>	0.88	1	1	0.96	0.5	1	1	0.96	0.54	1
<i>Rhopalodia</i>	0.39		0.15	0.05	0.13	0.5	0.5	0.04	0.15	
<i>Rossithidium</i>	0.02			0.07				0.07		
<i>Sellaphora</i>	0.23	0.75	0.37	0.31	0.5	0.5		0.21	0.08	
<i>Simonsenia</i>	0.2	0.25	0.11						0.23	0.5
<i>Staurosirella</i>	0.2	1	0.59	0.55	0.38	1	0.5	0.54	0.62	0.5
<i>Stauroneis</i>	0.04		0.04	0.22	0.13			0.07	0.15	
<i>Stauroforma</i>	0.05							0.04		
<i>Staurosira</i>	0.38	0.25	0.33	0.35	0.38	1	0.5	0.36	0.31	0.5
<i>Stephanodiscus</i>	0.12			0.02						
<i>Surirella</i>	0.18	0.25	0.15	0.2	0.13			0.18	0.38	1

Table 3 (con't)

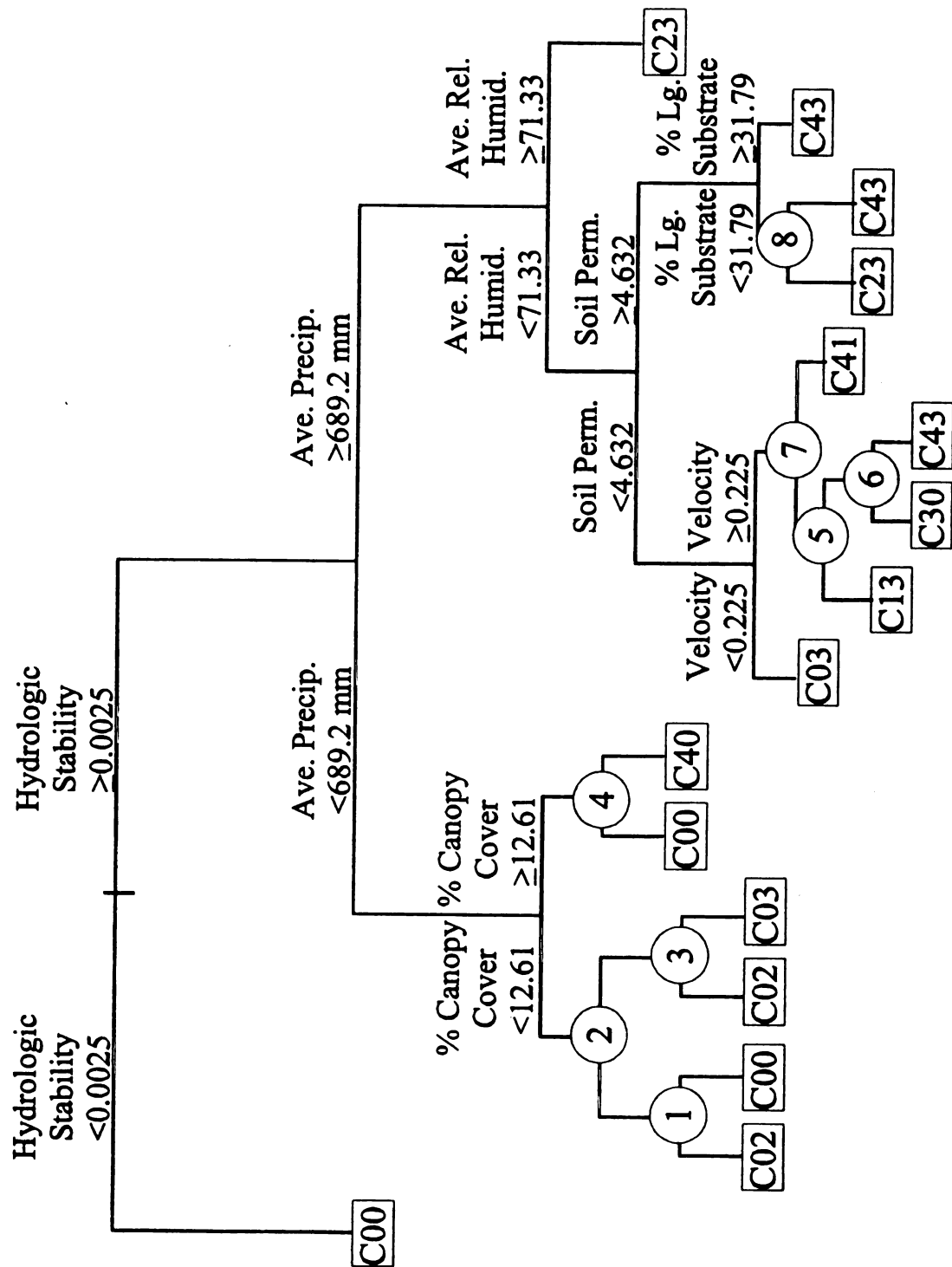
Genus	Proportion of Sites Per Class								
	C23	C30	C31	C32	C33	C40	C41	C42	C43
<i>Pseudostaurosira</i>	0.13	0.14	1	0.67	0.05	0.04	0.06		0.21
<i>Punctastriata</i>							0.03		
<i>Reimeria</i>	0.73	0.71	1	1	0.68	0.71	0.57	1	0.85
<i>Rhoicosphenia</i>	0.8	0.43		0.33	0.79	0.29	0.37		0.49
<i>Rhopalodia</i>		0.14				0.07			0.08
<i>Rossithidium</i>	0.13	0.07			0.05		0.2		0.19
<i>Sellaphora</i>	0.2	0.29			0.16	0.07	0.03		0.13
<i>Simonsenia</i>	0.03	0.21				0.14			0.02
<i>Staurosirella</i>	0.33	0.79	1	0.67	0.32	0.36	0.31	1	0.51
<i>Stauroneis</i>	0.2	0.14				0.07			0.04
<i>Stauroforma</i>							0.03		
<i>Staurosira</i>	0.3	0.36		0.33	0.32	0.11	0.14	1	0.49
<i>Stephanodiscus</i>	0.03								0.02
<i>Surirella</i>	0.2	0.29			0.11	0.11	0.09		0.09

monthly temperatures were negatively related to this axis. These variables also had large coefficients for the second and third axes. Elevation was also an important factor in the second discriminant axis.

CART Model Summary

A classification tree with 16 terminal nodes was selected because it had the smallest cross-validated relative error (Figure 12). Hydrologic stability and average annual precipitation were the two splits that explained the most variation in the data. Sites with a minimum mean monthly stream flow less than 0.25% of the maximum mean monthly stream flow were likely to belong to class C00; 70% of observations in this terminal node belonged to C00. Within streams that were more hydrologically stable, streams with less than 689.2 mm of average annual precipitation were different from those that had more precipitation. Streams with less precipitation were more likely to belong to C00 than those with more precipitation, which were more likely to belong to class C43. A fair amount of variation remained in these nodes, however. Streams with less than 689.2 mm of precipitation could be split into more homogeneous groups using percent canopy cover, maximum annual temperature, soil depth, longitude, and hydrologic stability. Streams with more

Figure 12. Classification tree (CART) for predicting diatom class membership using environmental variables. Decision rules at each split indicate which branch to take. Terminal nodes indicate predicted class membership. Decision rules that did not fit neatly onto the figure are indicated by circled numbers at the node and are as follows: 1) longitude ≥ -109.4 to the left, longitude < -109.4 to the right; 2) maximum temperature ≥ 13.23 to the left, maximum temperature < 13.23 to the right; 3) hydrologic stability ≥ 0.09905 to the left, hydrologic stability < 0.09905 to the right; 4) bedrock depth ≥ 46.54 to the left; bedrock depth < 46.54 to the right; 5) number of wet days < 98.7 to the left, number of wet days ≥ 98.7 to the right; 6) soil organic matter < 1.038 to the left, soil organic matter ≥ 1.038 to the right; 7) dominant geology granitic, mixed, sedimentary, or volcanic to the left, dominant geology gneiss or ultra-mafic to the right; 8) percent canopy cover ≥ 62.34 to the left, percent canopy cover < 62.34 to the right. Several variables are described in more detail in the text.



than 689.2 mm of precipitation could be split into more homogeneous groups using relative humidity, soil permeability, stream velocity, percent of substrata larger than 128 mm in diameter, dominant catchment geology, percent canopy cover, average number of days with measurable precipitation, and soil organic matter content. In total, the 15 splits in the tree explained roughly 36.5% of variation in the data, estimated following cross-validation.

Model Performance

The hybrid modeling approach outperformed LDA and CART models used independently (Table 4). The hybrid modeling approach correctly classified 216 sites, while LDA and CART correctly classified 179 and 201 sites, respectively. When factoring in probabilities associated with maximum likelihood estimates, improvement using the hybrid method was even more substantial. Within classes, LDA misclassifications ranged from 0% to 100% (Table 5). Low and high misclassification rates were most common in small classes. For classes that had more than 20 sites, LDA correctly classified anywhere from 18% to 76% of sites. The CART model often misclassified 100% of sites in small classes (Table 6). This is because little improvement in the model occurs by splitting these small groups from a

Table 4. Performance of linear discriminant analysis (LDA), recursive partitioning (CART), and the Bayesian hybrid predictive classification models with respect to misclassification rates, the odds of correctly classifying a site relative to the null model (random assignment to a class) and confidence-weighted ratios (CWR) of correct-to-incorrect classifications.

	LDA	CART	Hybrid
Percent Misclassified	56.17%	50.74%	47.06%
Odds > Chance	8.34	9.36	10.06
CWR	1.05	1.15	1.39

Table 5. Observed and predicted classifications using linear discriminant analysis (LDA).

Class Sites	Number of Sites	Predicted Class Membership																	Percent of Sites Correctly Predicted		
		C00	C01	C02	C03	C10	C11	C12	C13	C20	C22	C23	C30	C31	C32	C33	C40	C41		C42	C43
C00	82	62	8	5	1									1	1		1			3	75.61%
C01	4	1	2					1													50.00%
C02	27	8	11	4			1	1									1	1			40.74%
C03	55	5	4	29			1		3		1					1	2	2		7	52.73%
C10	8	3	1		2			1	1												25.00%
C11	2		1																	1	0.00%
C12	2	1					1														50.00%
C13	28	4	2	6				5		1	1	2					1			6	17.86%
C20	13	7		1					4			1									30.77%
C22	2			1						1											50.00%
C23	30	2	2	1			1	2			10						2	3	1	6	33.33%
C30	14	2		2					1			5					2			2	35.71%
C31	1													1							100.00%
C32	3	1		1							1										0.00%
C33	19		3	2				1			4					2				7	10.53%
C40	28	5		5					1							11		1		5	39.29%
C41	35	6	1	4				2			4						7			11	20.00%

Table 6. Observed and predicted classifications using the classification tree (CART).

Class Sites	Number of	Predicted Class Membership																Percent of Sites Corretly Predicted			
		C00	C01	C02	C03	C10	C11	C12	C13	C20	C22	C23	C30	C31	C32	C33	C40		C41	C42	C43
C00	82	72		3	4				1											2	87.80%
C01	4				2												1			1	0.00%
C02	27	6		15	3							1								2	55.56%
C03	55	4		3	37				1			2	1							7	67.27%
C10	8	4			1			1					1							1	0.00%
C11	2	1																		1	0.00%
C12	2				1							1									0.00%
C13	28	2		1	6			7			4							1		7	25.00%
C20	13	5			3								2				1			2	0.00%
C22	2				1															1	0.00%
C23	30	2			2				2			16					2			6	53.33%
C30	14	2			2			1					4				2			3	28.57%
C31	1	1																			0.00%
C32	3				2							1									0.00%
C33	19	2		1	5							1						2		8	0.00%
C40	28	6		1	9												6			6	21.43%
C41	35	1		1	5				1			11						4		12	11.43%

larger group. Sites in larger classes were correctly classified between 11% and 89% of the time. Improvements in correct classification rates by CART over LDA were due to lower rates of misclassification in larger classes. The hybrid method was unable to eliminate problems associated with misclassification in small classes (Table 7). Misclassification rates for these small classes ranged from 0% to 100%, with 0% being far more common. Sites in larger classes were correctly classified between 29% and 88% of the time. Overall model performance with respect to the indices used here was highest for the hybrid method and lowest for the standard LDA method, while CART performance was intermediate.

Table 7. Observed and posterior predicted classifications using the hybrid method combining linear discriminant analysis (LDA) and classification tree (CART) predictions.

Class Sites	Number of Sites	Predicted Class Membership																Percent of Sites Correctly Predicted		
		C00	C01	C02	C03	C10	C11	C12	C13	C20	C22	C23	C30	C31	C32	C33	C40		C41	C42
C00	82	72		3	4				1										2	87.80%
C01	4				2											1			1	0.00%
C02	27	6		15	3							1					1		1	55.56%
C03	55	2		3	37			1	1	2		1	1			1			6	67.27%
C10	8	3			1	1		1	1	1									1	12.50%
C11	2			1															1	0.00%
C12	2							1				1								50.00%
C13	28	1		1	6				8			4					1		7	28.57%
C20	13	4			2					3			1			1			2	23.08%
C22	2				1														1	0.00%
C23	30	1			3				2			13				3	3		5	43.33%
C30	14	2			4				1	1			2			2			2	14.29%
C31	1													1						100.00%
C32	3				2							1								0.00%
C33	19			2	4				1			1			2	1	2		6	10.53%
C40	28	3			8					1						10			6	35.71%
C41	35				5				1			6					12		11	34.29%

DISCUSSION

Predictive models are valuable tools that are increasingly being applied in ecological assessment and environmental risk assessment (Côté and Reynolds 2002). Current methods being used in predictive classification have different strengths and weaknesses. The Bayesian approach presented here provides a formal method to integrate the predictions of linear and nonlinear predictive classification models. Linear discriminant analysis and recursive partitioning provide alternative methods for developing predictive classification models. Each method has its own benefits and problems but the two methods can be used in a complimentary way using Bayes' theorem as described.

The hybrid approach used here resulted in lower misclassification rates than those observed for either LDA or CART alone. Furthermore, the CWR suggested that improvement in inference using the hybrid method was even higher than suggested by misclassification rate. In RIVPACS-type models, probabilistic classifications are used rather than maximum likelihood. Therefore, metrics such as the CWR which incorporate classification probabilities are more representative of the improvement likely to be observed by combining linear and nonlinear models in a

formal way. The hybrid method resulted in increased classification probabilities when LDA and CART both suggested that a site had a high probability of belonging to a given class. When the two modeling approaches exhibited significant differences in classification probabilities, the hybrid method provided a quantitative intermediate to these predictions, which can be important when the true class of a given observation is not known. In fact, if the true class membership were known, there would be no need for these types of predictive models in environmental assessment. Previous qualitative syntheses of modeling results give equal weight to each maximum likelihood prediction, despite the knowledge that one of these predictions is wrong.

In addition to improved inference, using a combination of linear and nonlinear models can provide more complete insight to ecological patterns because both linear and nonlinear patterns are likely in ecological systems. Linear discriminant analysis suggested that temperature was the most important factor for discriminating stream classes based on diatom genus composition. While this analysis is observational in nature, thereby making cause-and-effect relationships difficult to conclusively establish, previous research does suggest that temperature may be an important factor regulating the composition of algal communities in

streams. Many studies have shown observational relationships between temperature and algal taxonomic composition in streams (Pan et al. 1999, Lamberti and Resh 1985, Wilde 1982, Squires et al. 1979), but experiments establishing causal relationships are rare (Wilde and Tilly 1981). In general, these studies show a shift in periphyton composition from diatoms at low temperatures, to chlorophytes and xanthophytes at intermediate temperatures, and cyanobacteria at higher temperatures. Diatom assemblages themselves also change along temperature gradients (Potapova and Charles 2002, Vinson and Rushforth 1989, Squires et al. 1979, Klarer and Hickman 1975, Patrick 1971), with increases in the genus *Cocconeis* commonly observed. While causal relationships with temperature may, in part, be responsible for the observed patterns in diatom genus composition, unobserved covariates are possible when observing patterns at such large spatial scales.

The most important nonlinear patterns observed in the CART model were hydrological in nature. The variable explaining most of the variability in diatom genus class was the hydrologic stability index. The first split in the tree isolated a significant proportion of streams in the diatom class C00. This was the largest class, containing 20% of all sites. This split separated a group of sites characterized by an annual minimum monthly discharge that

was just a fraction of the annual maximum monthly discharge. These streams appear to be characterized by small baseflow and large event flow; in other words, these streams are not maintained by persistent sources such as ground-water circulation and have months of high input such as that supplied by snow melt. In fact, many of the streams in this class were found in North Dakota and South Dakota where snowfall is high and significant aquifer input is unlikely (McGuinness 1963). Others have found that these hydrologic characteristics can influence water chemistry during the summer, which may have an influence on stream periphyton (Andersen et al. 2005, Peterson et al. 2001). This class of streams was characterized by higher conductivity and total phosphorus than streams at the opposite side of the SOM, suggesting that hydrologic regulation of water chemistry may be regulating diatom genus composition.

Recent applications of predictive classification models such as RIVPACS attempt to quantify the expected biological community at a given site in the absence (or minimization) of anthropogenic disturbance (Moss et al. 1987) in order to assess the biological integrity of the stream. Significant differences between observed and expected assemblages indicate ecological impairment. One potential problem with current RIVPACS-type models is the sole use of LDA to

develop predictive classification models. Ecological data often violate several of the assumptions of LDA and missing data are common in most bioassessment surveys. The hybrid approach presented here will result in more reliable predictive classification models for RIVPACS-type assessments and may improve inference in other ecological classification problems, such as species profiling for invasiveness risk.

REFERENCES

- Andersen, H. E., Kronvang, B., and Larsen, S. E. 2005. Development, validation and application of Danish empirical phosphorus models. *Journal of Hydrology* 304: 355-365.
- Barbour, M. T., Gerritsen, J., Griffith, G. E., Frydenborg, R., McCarron, E., White, J. S., and Bastian, M. L. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15: 185-211.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. 1984. *Classification and Regression Trees*, Chapman and Hall/CRC, Boca Raton, FL.
- Clarke, R. T., Wright, J. F. and Furse, M. T. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling* 160: 219-233.
- Côté, I. M. and Reynolds, J. D. 2002. Predictive ecology to the rescue? *Science* 298: 1181-1182.
- DeShon, J. E. 1995. Development and application of the invertebrate community index (ICI). Pages 217-243 in *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Davis, W. S., and Simon, T. P., eds. Lewis Publishers, Boca Raton, FL.
- Gevrey, M., Rimet, F., Park, Y. S., Giraudel, J. L., Ector, L., and Lek, S. 2004. Water quality assessment using diatom assemblages and advanced modelling techniques. *Freshwater Biology* 49: 208-220.
- Guo, Q. F. and Brown, J. H. 1996. Temporal fluctuations and experimental effects in desert plant communities. *Oecologia* 107: 568-577.
- Hansmann, E. E., and Phinney, H. K. 1973. Effects of logging on periphyton in coastal streams of Oregon. *Ecology* 54: 194-199.

- Hawkins, C. P., Norris, R. H., Hogue, J. N. and Feminella, J. W. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10: 1456-1477.
- Hill, M. O. 1979. TWINSpan-A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and the attributes. *Ecology and Systematics*, Cornell University, Ithaca, NY. 52 pp.
- Klarer, D. M., and Hickman, M. 1975. The effect of thermal effluent upon the standing crop of an epiphytic algal community. *Hydrobiologia* 60: 17-62.
- Kolar, C. S. and Lodge, D. M. 2002. Ecological predictions and risk assessment for alien fishes in North America. *Science* 298: 1233-1236.
- Lamberti, G. A., and Resh, V. H. 1985. Distribution of benthic algae and macroinvertebrates along a thermal stream gradient. *Hydrobiologia* 128: 13-21.
- Lek, S., Guirèsse, M., and Giraudel, J. L. Predicting stream nitrogen concentration from watershed features using neural networks. *Water Research* 33: 3469-3478.
- Lowe, R. L., and Pan, Y. 1996. Benthic algal communities as biological monitors. Pages 705-739 in *Algal Ecology: Freshwater Benthic Ecosystems*, Stevenson, R. J., Bothwell, M. L., and Lowe, R. L., eds. Academic Press, San Diego, CA.
- McCune, B. and Grace, J. B. 2002. *Analysis of Ecological Communities*, MJM Software Design, Glenden Beach, OR.
- McGuinness, C. L. 1963. The role of ground water in the national water situation. Washington, D.C. United States Geological Survey. Water-Supply Paper 1800.
- Moss, D. 2000. Evolution of statistical methods in RIVPACS. In *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*, J. F. Wright, D. W. Sutcliffe, M. T. and Furse (eds), Freshwater Biological Association, Ambleside, Cumbria, UK. pp. 25-37.

- Moss, D., Furse, M. T., Wright, J. F. and Armitage, P. D. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 17: 41-52.
- Norris, R. H. and Norris, K. R. 1995. The need for biological assessment of water quality: Australian perspective. *Australian Journal of Ecology* 20: 1-6.
- Ohio Environmental Protection Agency (Ohio EPA). 1987. Biological criteria for the protection of aquatic life, Volumes I-III. Ohio EPA, Columbus, OH.
- Ostermiller, J. D., and Hawkins, C. P. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. *Journal of the North American Benthological Society* 23: 363-382.
- Pan, Y., Stevenson, R. J., Hill, B. H., Kaufmann, P. R., and Herlihy, A. T. 1999. Spatial patterns and ecological determinants of benthic algal assemblages in Mid-Atlantic streams, USA. *Journal of Phycology* 35: 460-468.
- Patrick, R. 1971. The effects of increasing light and temperature on the structure of diatom communities. *Limnology and Oceanography* 16: 405-421.
- Peterson, C. G., Valett, H. M., and Dahm, C. N. 2001. Shifts in habitat templates for lotic microalgae linked to interannual variation in snowmelt intensity. *Limnology and Oceanography* 46: 858-870.
- Potapova, M. G, and Charles, D. F. 2002. Benthic diatoms in USA rivers: distributions along spatial and environmental gradients. *Journal of Biogeography* 29: 167-187.
- Recknagel, F., French, M., Harkonen, P., and Yabunaka, K. I. 1997. Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling* 96: 11-28.
- R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

- Reichard, S. H. and Hamilton, C. W. 1997. Predicting invasions of woody plants introduced into North America. *Conservation Biology* 11: 193-203.
- Rejmanek, M. and Richardson, D. M. 1996. What attributes make some plant species more invasive? *Ecology* 77: 1655-1661.
- Reynoldson, T. B., Bailey, R. C., Day, K. E. and Norris, R. H. 1995. Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology* 20: 198-219.
- Squires, L. E., Rushforth, S. R., and Brotherson, D. J. 1979. Algal response to a thermal effluent: Study of a power station on the Provo River, Utah, USA. *Hydrobiologia* 63: 1011-1017.
- Wilde, E. W. 1982. Responses of attached algal communities to termination of thermal pollution. *Hydrobiologia* 94: 135-138.
- Wilde, E. W., and Tilly, L. J. 1981. Structural characteristics of algal communities in thermally altered artificial streams. *Hydrobiologia* 76: 57-63.
- Williams, B. K. 1983. Some observations of the use of discriminant analysis in ecology. *Ecology* 64: 1283-1291.
- Yoder, C. O., and Rankin, E. T. 1995. Biological criteria program development and implementation in Ohio. Pages 109-144 in *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Davis, W. S., and Simon, T. P., eds. Lewis Publishers, Boca Raton, FL.

CHAPTER FOUR

A RELATIVE RISK FRAMEWORK FOR DEVELOPING NUTRIENT WATER QUALITY CRITERIA BY INFERRING REFERENCE CONDITIONS AND THE PROBABILITY OF LOSING VALUED ECOLOGICAL ATTRIBUTES

INTRODUCTION

Cultural eutrophication is one of the most widespread threats to water quality in the United States. According to a report by the United States Environmental Protection Agency (USEPA), 40% of surveyed streams and rivers in the nation are impaired due to nitrogen and phosphorus enrichment (USEPA 1996). As a result, the Clean Water Action Plan was released (USEPA 1998), directing the USEPA to develop and publish nutrient criteria guidance to assist states and tribes in developing water quality standards. In 2000, the USEPA published nutrient criteria guidance for streams and rivers (USEPA 2000) prescribing a combination of approaches including classification, stressor-response relationships, establishing reference conditions, and using published research.

Developing water quality criteria for nutrients is particularly challenging because nutrients are naturally present in streams and rivers, nutrient concentrations vary across stream types, and nutrients are necessary for the normal functioning of biological systems. Effective nutrient criteria must account for this natural variability in stream nutrient concentrations, while protecting biological integrity. USEPA guidance recommends classifying streams to help account for variation in

nutrient concentrations due to natural factors such as geology, hydrology, and climate. In the development of its nutrient criteria recommendations, the USEPA applies an ecoregion approach to account for natural variation in N and P. The problem with this regionalization approach is that it assumes all streams within an ecoregion are similar and it can be confounded with differences in human disturbance among regions. This can lead to artificially high nutrient criteria in developed regions or artificially low criteria for streams with naturally high nutrient concentrations due to geology. Alternative approaches presented in the guidance document include classifications by stream order, geomorphology, and geology. One important aspect of classification that is absent from USEPA recommendations is that classifications should be relatively independent of human activities to avoid potentially confounding sources of natural variation in nutrients.

Reference site approaches have been advocated as a way to establish water quality criteria by characterizing background levels in nitrogen and phosphorus. This approach generally uses a sample of streams with minimal human disturbance to characterize what streams should look like under minimally disturbed conditions. This approach to characterizing expected conditions can be difficult

because human activities have affected nutrient concentrations in most streams and rivers, leaving relatively few minimally disturbed sites or restricting reference sites to a limited geographic region.

Predictive models have recently been applied to overcome a lack of adequate reference sites. These models infer nutrient concentrations in the absence of human disturbance (Dodds and Oakes 2004, Smith et al. 2003) by constructing statistical models that predict nutrient concentrations using natural and anthropogenic variables; anthropogenic factors are then removed and nutrient concentrations in the absence of human disturbance can be inferred. Smith et al. (2003) used this approach to correct for atmospheric deposition of nutrients at reference sites, while Dodds and Oakes (2004) applied this approach to remove the effects of agricultural and urban land use in a set of sites varying in human disturbance.

In addition to potential difficulties finding adequate sites with minimal human disturbance, the reference-based approach has also been criticized because reference-derived nutrient criteria may be over-protective and unrepresentative of societal values (Reckhow et al. 2005). Water quality criteria are established to protect designated uses and should therefore be good surrogates for the attainment of those designated uses. Therefore,

Reckhow and others advocate an effects-based approach that uses the probability of attaining designated uses as the endpoint. Case studies presented by Reckhow et al. rely heavily on best professional judgment and are restricted to single water bodies; however, they envision regional application of the method, as well as incorporating user response surveys to quantify the probability of attaining designated uses more rigorously.

Effects-based criteria have also been suggested by others, although biological attributes are generally suggested as endpoints, rather than designated use attainment. This approach is often used for toxic chemicals. Generally, criteria for toxic chemicals are established at a level well below a threshold above which individual organisms are likely to die in laboratory tests. This individual-based dose-response approach has been criticized by those in favor of more ecologically relevant endpoints (Cairns and Pratt 1986, Cairns 1983). Others have proposed stressor-response thresholds in more realistic systems may be useful for establishing water quality criteria (Stevenson et al. 2004, King and Richardson 2003). These thresholds can be good indicators that human activities have affected water quality in aquatic ecosystems and can be used to establish benchmarks for nutrient criteria.

Reference-based approaches and stressor-response relationships address separate questions in the development of nutrient criteria. These questions include (1) at what concentration does a nutrient have undesired effects on valued biological characteristics of a stream? (2) what should a system look like in the absence of human disturbance or at an acceptable level of human disturbance? (3) assuming that a stream is in reference condition, what is the probability that a given valued biological characteristic would be supported?

Reference approaches and stressor-response thresholds are generally presented as alternative ways in which water quality criteria can be developed (Stevenson et al. 2004, USEPA 2000). Although both approaches are recommended for informing nutrient criteria development, little guidance exists for integrating the results of these two approaches. Currently, an effective framework for nutrient criteria development that integrates reference-based approaches, effects-based approaches, and societal goals is lacking. In this paper, I present a relative risk (RR) framework for setting site-specific benchmarks representing candidate nutrient criteria. This framework formally incorporates findings of previous research, effects-based endpoints, and a reference-based approach meant to reflect societal goals for water quality standards using Bayesian statistical

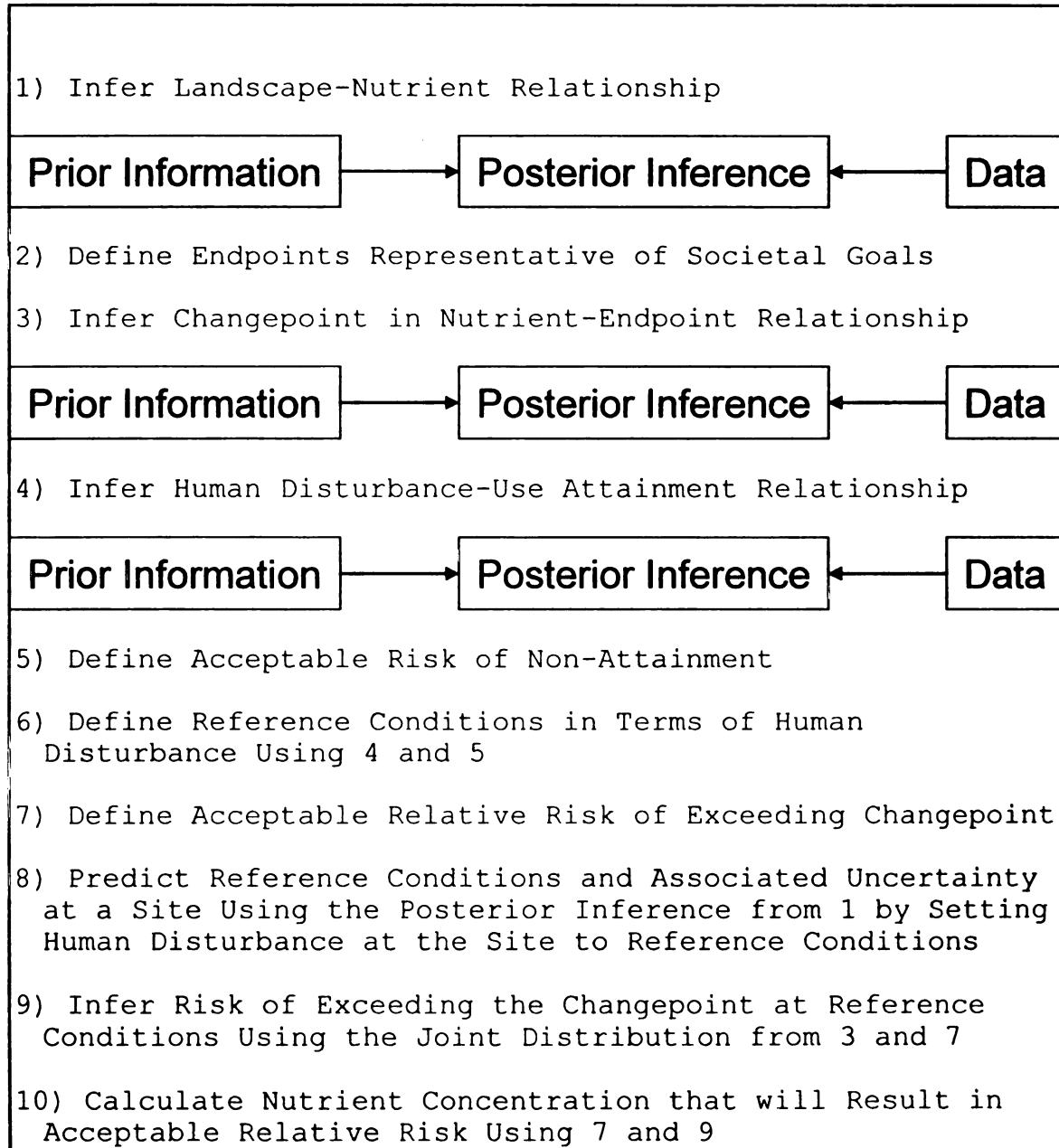
methods. The approach is applied to Michigan streams with emphasis on total phosphorus (mg/L TP) and its effects on water column chlorophyll (mg/L fluorometric, phaeopigment-corrected chlorophyll a) using data from the National Nutrient Database which is available through the USEPA (www.epa.gov/waterscience/criteria/nutrient/). For the state of Michigan, the database is primarily populated by sites sampled by the Michigan Department of Environmental Quality (MDEQ) and the United States Geological Survey. The data include information from streams and rivers; impoundments are commonly classified as lakes by MDEQ, which were excluded from this analysis.

Bayesian inference is well suited for integrating multiple sources of information for environmental decision-making under a risk framework. Several authors have highlighted the advantages of Bayesian approaches for addressing environmental problems and in communicating results (Carpenter 2002, Wade 2000, Ellison 1996, Reckhow 1990). Among the benefits of Bayesian statistics is the ability to incorporate information from previous research. Unlike classical statistics, specification of prior parameter estimates is fundamental to the Bayesian approach. Another benefit of the Bayesian method is the explicit quantification and propagation of uncertainty in complex models. Environmental management involves

decision-making under conditions of uncertainty. Downplaying or failing to present this uncertainty in an easily interpretable manner can lead to poor management decisions (Peterson et al. 2003, Pielke and Conant 2003). The need to incorporate uncertainty in ecological prediction is increasingly voiced by ecologists (Brewer and Gross 2003, Clark 2003, Pielke and Conant 2003, Carpenter 2002).

I apply a Bayesian approach to inform nutrient criteria development, integrating prior information and explicitly incorporating prediction uncertainty. Stressor-response relationships, inferred reference conditions with minimal human disturbance, and relationships published in the literature are combined using a series of models. Bayesian methods are used to quantify the threshold relationship between total phosphorus and chlorophyll observed in Michigan streams and rivers and to infer expected levels of total phosphorus if human disturbance were minimized. Finally, benchmarks for TP criteria are determined using a relative risk approach that balances the current risk of exceeding the TP-chlorophyll threshold with the risk of surpassing the threshold at minimal levels of human disturbance. The steps involved in this framework are outlined in Figure 13, which precedes a more detailed descriptions of framework in the methods.

Figure 13. Steps involved in developing candidate nutrient criteria using the relative risk framework presented in this document. Prior information boxes represent understanding of a given relationship prior to analyzing the data. Priors include an estimate of central tendency, as well as uncertainty associated with the prior information. Posterior inference boxes represent the results of Bayesian analysis that integrates prior information and data.



METHODS

Environmental Thresholds

Thresholds along the total phosphorus gradient that resulted in significant changes in water column chlorophyll were quantified using a hierarchical Bayesian changepoint method described by Qian et al. (2003). Chlorophyll concentrations may change in mean and/or variance along a TP gradient. A threshold or changepoint is a value of TP along the ordered gradient that separates chlorophyll measurements into the two groups with the greatest difference in mean and/or variance. Under the Bayesian method, $y_1, \dots, y_n$ represents the sequence of observed chlorophyll concentrations along a sequence of ordered TP concentrations, $x_1, \dots, x_n$ at n sites. Observed chlorophyll concentration y_i is assumed to be drawn from a log-normal distribution, Y_i . An environmental threshold at r , where $1 \leq r \leq n$, separates two groups of chlorophyll variables: $Y_1, \dots, Y_r$ and $Y_{r+1}, \dots, Y_n$ such that

$$Y_i \sim \begin{cases} \log\text{-Normal}(\mu_1, \sigma_1^2), i=1, \dots, r \\ \log\text{-Normal}(\mu_2, \sigma_2^2), i=r+1, \dots, n \end{cases} \quad (8)$$

where μ_1 and σ_1^2 are the mean and variance, respectively, of chlorophyll concentration below the TP threshold; μ_2 and σ_2^2 are the mean and variance, respectively, of chlorophyll concentration above the TP threshold. Therefore, the TP threshold is a function of the mean and variance of chlorophyll observed above and below the threshold.

Bayesian analysis incorporates prior information into the analysis, making it unique from classical statistics which attempts "objectivity" by assuming no prior information (e.g., null hypothesis testing). Bayesian analysis acknowledges and utilizes prior information. Prior information is incorporated in an analysis using Bayes' theorem. The Bayesian approach results in a posterior model that combines the probability of prior information (the *prior*) and the probability of any set of random data given a specific model (the *likelihood*). The likelihood treats model parameters as fixed values and data as random, while the posterior treats model parameters as random and data are fixed. In other words, Bayesian analysis acknowledges uncertainty associated with the parameters and assumes the data are known. As a result, Bayesian model parameters such as slopes and intercepts in regression models are distributions, rather than fixed values.

Slightly more detailed treatments of Bayesian statistics accessible to ecologists are available (Bolstad 2004, Gotelli and Ellison 2004) and mathematical details for the Bayesian changepoint analysis applied in ecology has been described elsewhere (Qian et al. 2003). Because Bayesian analyses requires that prior information about a model be specified in a probabilistic way (i.e., the central tendency along with associated uncertainty), distributions for prior information on the chlorophyll means, variances, and changepoint need to be specified. The prior distribution for the changepoint was assumed to take a gamma distribution, with a mean of 0.030 mg/L TP and a variance of 300 by setting the shape parameter, α , equal to 3 and the scale parameter, β , equal to 10. The gamma, much like the normal, is a function describing the central tendency and variance of a variable. The gamma distribution can accommodate skew and cannot take values less than zero. The mean for the changepoint prior was chosen following the results of Dodds et al. (2002), who reported a parametric breakpoint for chlorophyll along a TP gradient at 0.031 mg/L TP and a non-parametric breakpoint at 0.029 mg/L TP. The variance was chosen such that the prior would be relatively broad, but not completely uninformative. Published data on water column chlorophyll in streams is somewhat limited. Prior means were

determined using approximations modeled from median TP values from the data, above and below the mean of the prior threshold (0.010 mg/L and 0.080 mg/L TP, respectively) using a model published by Van Nieuwenhuyse and Jones (1996)

$$\log Chl = -1.65 + 1.99(\log TP) - 0.28(\log TP)^2, \quad (9)$$

$s=0.32$, $R^2=0.67$, $F=291$, $p<0.001$, and $n=292$, where chlorophyll and TP concentrations are both in $\mu\text{g/L}$. This approach suggested a mean chlorophyll concentration below the threshold of approximately 0.0012 mg/L and a mean above the threshold of approximately 0.0133 mg/L. One hundred was used as the variance for the prior chlorophyll distribution below the threshold and 1000 was used for the variance above the threshold, making the prior distributions weak and only somewhat informative.

Complex Bayesian models can be difficult or even impossible to calculate analytically. Recent improvements in computers have made computationally intensive methods available that are capable of providing estimates for integrals of high-dimensional probability distributions. Markov Chain Monte Carlo (MCMC) is now commonly applied in Bayesian analyses (Gilks et al. 1996). Several programs are available for conducting MCMC, including the free software WinBUGS, developed as part of the BUGS (Bayesian

inference Using Gibbs Sampling) project (Spiegelhalter et al. 2004). WinBUGS implements one type of MCMC algorithm called the Gibbs sampler. A Gibbs sampling algorithm was used for both the changepoint analysis and the predictive model described in the next session.

Predictive Models

Predictive models were constructed to infer TP concentrations if human disturbance were minimized. Total phosphorus was modeled using Bayesian multiple linear regression. Predictor variables were selected from a variety of stream characteristics including channel length, sinuosity, and gradient, as well as watershed characteristics such as precipitation, slope, urban and crop land use. The Julian day on which samples were collected was also evaluated as a potential variable. These variables were selected following removal of redundant, highly correlated, and/or sparse variables from a larger set potential variables. The best subset of remaining variables was then chosen using stepwise forward and backward selection based on Akaike's Information Criterion (AIC; Akaike 1973, 1985). AIC is an optimization metric that is minimized when the likelihood of the data, given the model, is maximized and the number of parameters is minimized. AIC adds a greater penalty to models that

use more parameters. The best model has the lowest AIC. The final set of predictor variables included in the model were natural log-transformed channel length (i.e., a hydrologically similar stream reach, such as that found between two points of confluence or between a lake and a point of confluence), natural log-transformed precipitation, percent urban land use, and percent crop land use.

Normal distributions were used to describe priors for the regression slope parameters. Diffuse, uninformative priors were used for channel length and precipitation parameters. The means for these priors were set to zero and variances of 1000 were used. Informed priors were used for the percent urban and percent crop parameters. These priors were defined using the model published by Dodds and Oakes (2004). The percent crop prior was centered at 0.00668 with a standard error of 0.001097. The prior for the percent urban slope parameter was given a mean of 0.01465 and a standard error of 0.003280. The conditional error variance was given a diffuse inverse-gamma prior with shape and rate parameters set to 0.001. Details for semi-conjugate priors in regression models can be found elsewhere (e.g., Gelman et al. 1995).

Posterior inference using Bayesian regression models incorporates uncertainty in the model, as well as sampling

error. The posterior variance is equal to the variance of the model plus the error variance. Predictive distributions for an unobserved response, $\tilde{y}$, given a new set of observations, $\tilde{X}$, can be simulated by repeatedly drawing from the joint posterior distribution of the model parameters, inputting $\tilde{X}$ to the equation, and adding random draws from a normal distribution with a mean of zero and a variance equal to a random draw from the posterior error variance distribution. Using this simulation method, retrospective predictions of reference TP and its uncertainty can be estimated using the channel length and watershed precipitation for a site, and the percent of urban and crop land use consistent with the definition of reference conditions described in the next section. These four variables are the elements of $\tilde{X}$.

Defining Reference Conditions

Reference conditions reflective of designated use attainment can be established using logistic regression models that predict the probability of designated use attainment as functions of agricultural and urban land use. However, logistic regression has a fixed parametric form that can distort the land use-attainment relationship if the data do not take this form. This was the case with the Michigan data, where observations in the middle and upper

range of land use values caused obvious underestimates in attainment probabilities at the low end of the gradient. This was particularly evident in the urban gradient shown in Figure 14a. The logistic model provided a poor fit to these data and suggested the probability of non-attainment at zero urbanization was greater than 20%.

Locally weighted regression (LOESS) can also be used to depict land use-attainment relationships. The LOESS smoother predicts values for the response variable using weighted least squares of the predictor variable and $k\%$ nearest-neighbors, with higher weights given to closer neighbors (Hastie and Tibshirani 1990). Using this method, attainment probabilities can be modelled across the land use gradient providing better local predictions when logistic models are ineffective at explaining patterns in the data. Graphical model presentation can then be used to estimate percents of urban and crop land use at which attainment probability is high (Figure 14).

Establishing Benchmarks Using Relative Risk

Relative risk measures the influence of a risk factor on a specified outcome. Relative risk is a concept commonly applied in epidemiology as the ratio of the incidence rate of a disease among individuals exposed to a risk factor to the incidence rate of individuals not

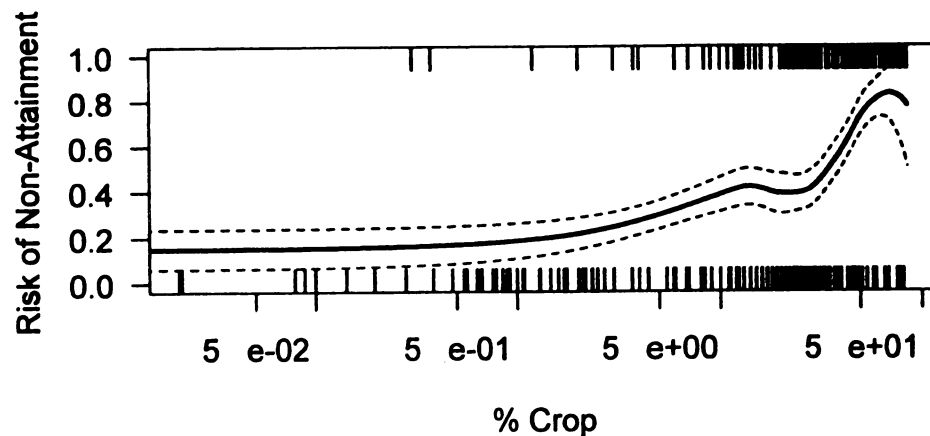
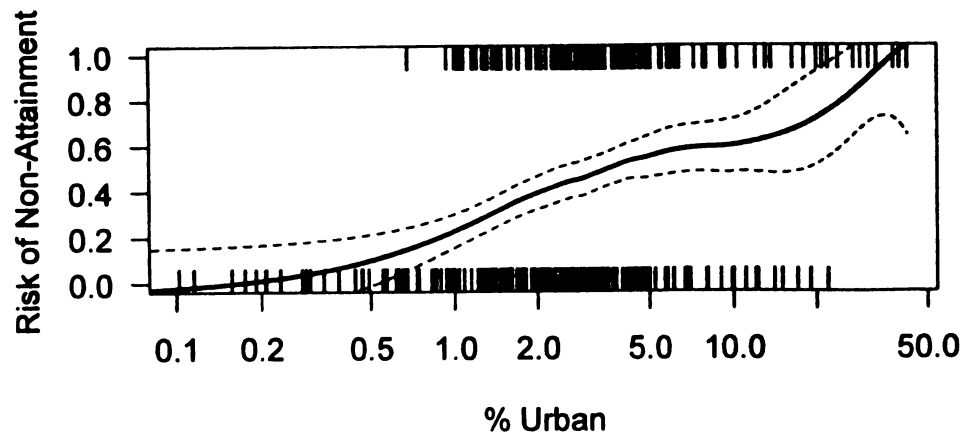


Figure 14. Probability of not attaining Category 2 status from the Michigan Department of Environmental Quality 305(b) report as a function of percent urban (above) and percent crop (below) land use in the watershed. Curves represent LOESS fits (solid line) and pointwise 95% confidence intervals (dashed lines). Figure 1b is equivalent to figure 1a, with the exception of a log-transformed x-axis to improve resolution at the low end of the gradient.

exposed to the risk factor. For example, the probability of lung cancer for a smoker divided by the probability of lung cancer for a non-smoker represents the relative risk of lung cancer for smokers. The relative risk concept can be applied to environmental management to compare the likely causes of an undesired outcome, such as population decline in fish (Landis et al. 2004). The relative risk model might also be used to compare the probability of some adverse event under different management options. Here I use RR to (1) examine the probability of exceeding the TP threshold at current TP levels, relative to the probability of exceeding the threshold at inferred reference levels, and (2) set benchmarks that provide candidate TP criteria. The former task is accomplished by determining the TP concentration at which $RR=1$. Alternative levels of RR might be chosen based on acceptable levels of risk. At $RR=1$, the probability of exceeding the threshold at the TP benchmark is equal to the probability that inferred reference conditions exceed the threshold. The probability that reference conditions exceed the threshold is equal to the joint probability of inferred reference TP and threshold TP (Figure 15, page 113). The probability that the current nutrient concentration exceeds the stressor-response threshold is equal to the proportion of the stressor response distribution that is below the current

nutrient concentration (Figure 15, page 114). The relative risk of current conditions over inferred reference conditions is equal to the probability that current nutrient concentrations exceed the threshold divided by the probability that inferred reference concentrations exceed the threshold. A site-specific nutrient benchmark value can be calculated at a $RR=1$, which represents a concentration at which risk of exceeding the threshold is equal to the risk under reference conditions. I present detailed results of this analysis for a site on the Cass River in Saginaw County, Michigan, USA, that has chlorophyll concentrations at the higher end of those observed in the dataset, 0.050 mg/L chlorophyll. Characteristics of this site include a channel length of 4086 m, watershed precipitation of 758.8 mm, 4% urban land use in the watershed, 56% crop land use in the watershed, and 0.090 mg/L TP.

In some cases, the probability that inferred reference nutrient concentrations are higher than the TP threshold may approach one. In these cases it is probably unreasonable to expect a site to support the valued biological attribute being considered and less sensitive endpoints should be evaluated. A threshold separating two states of water column chlorophyll concentrations in streams is the endpoint considered here. Some streams may

have inferred reference TP higher than the threshold. It would not make sense to set TP criteria for the site at a level intended to protect an attribute that has a low probability of existence to begin with. Rather, thresholds based on indirect effects such as invertebrate metrics, fish metrics, or dissolved oxygen might be used.

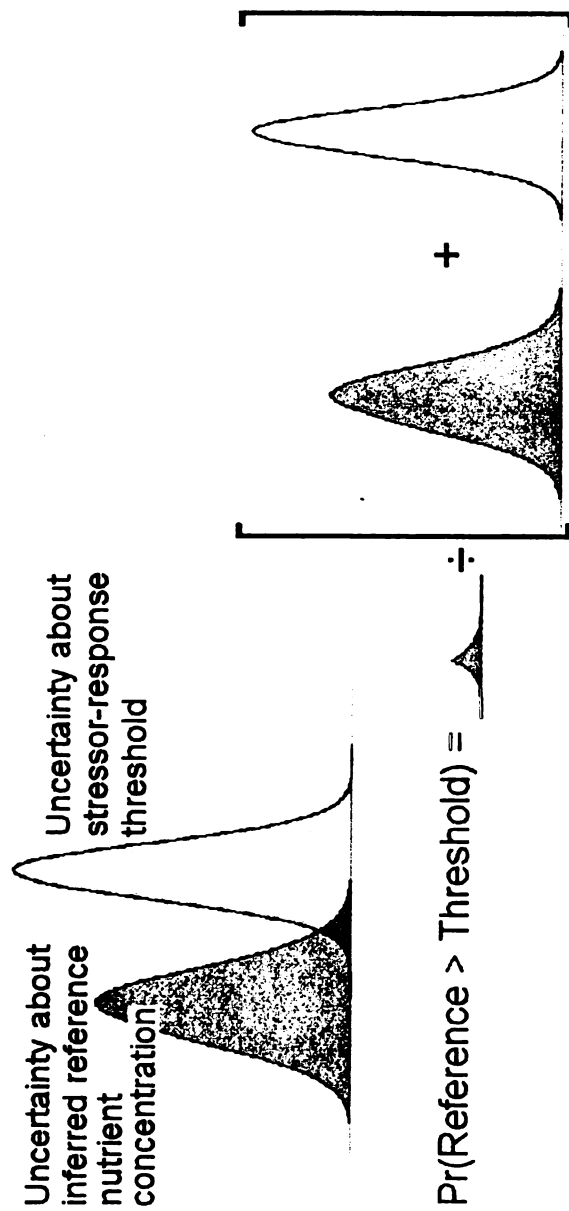
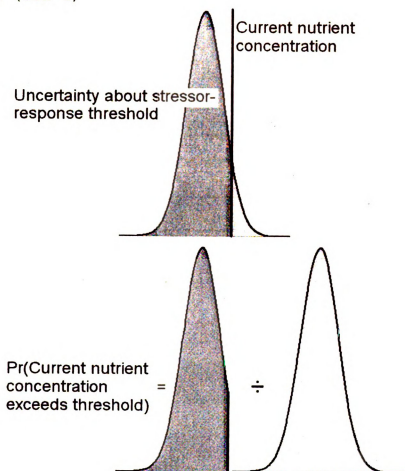


Figure 15. Graphical representation for calculating the probability that inferred reference nutrient concentrations are higher than the stressor-response threshold and the probability that the current nutrient concentration exceeds the stressor-response threshold (next page).

Figure 15 (con't)



RESULTS

The posterior TP changepoint distribution separating groups of streams characterized by low mean chlorophyll levels and streams characterized by high mean chlorophyll levels was concentrated in the region between 0.040 and 0.055 mg/L TP, with a median at 0.044 mg/L TP. Ninety-five percent of the posterior mean chlorophyll density below the threshold was between 0.0010 and 0.0017 mg/L chlorophyll *a*. Mean chlorophyll above the threshold was roughly 6 times higher than chlorophyll below the threshold.

Likelihood functions had a stronger influence on the TP changepoint and mean chlorophyll posterior distributions than the prior distributions, but the combination of prior information and the data decreased uncertainty in the parameter distributions. The posterior resulted in a higher changepoint than suggested by prior information and the width of the distribution was narrower than suggested by the likelihood distribution (Figure 16). Substantial improvement in posterior inference over the likelihoods, even with relatively weak prior information, was observed for mean chlorophyll below (Figure 17) and above (Figure 18) the changepoint.

Posterior estimates of the predictive model parameters mediated percent urban and crop parameters suggested by

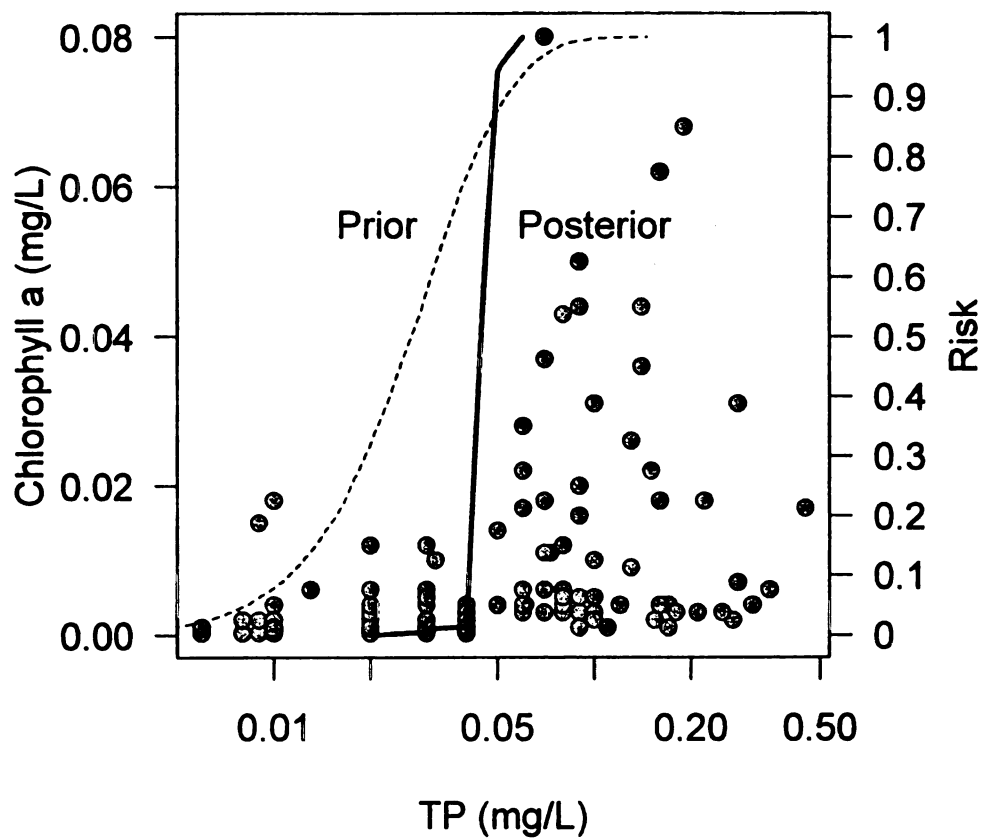


Figure 16. Total phosphorus-water column chlorophyll threshold. Cumulative frequency distributions of the prior changepoint (dashed line) and posterior changepoint (solid line) densities indicate risk of exceeding the changepoint separating a state of low mean chlorophyll and high mean chlorophyll.

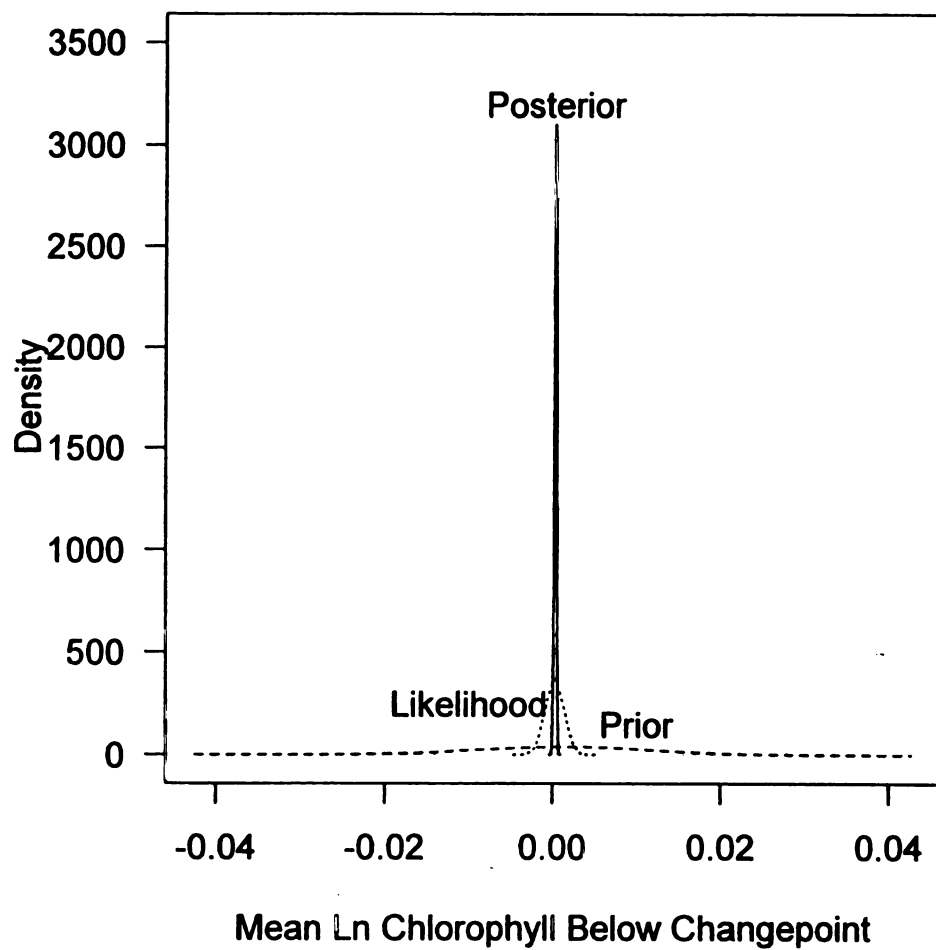


Figure 17. Prior, likelihood, and posterior densities of natural-log-transformed chlorophyll (mg/L) below the total phosphorus threshold.

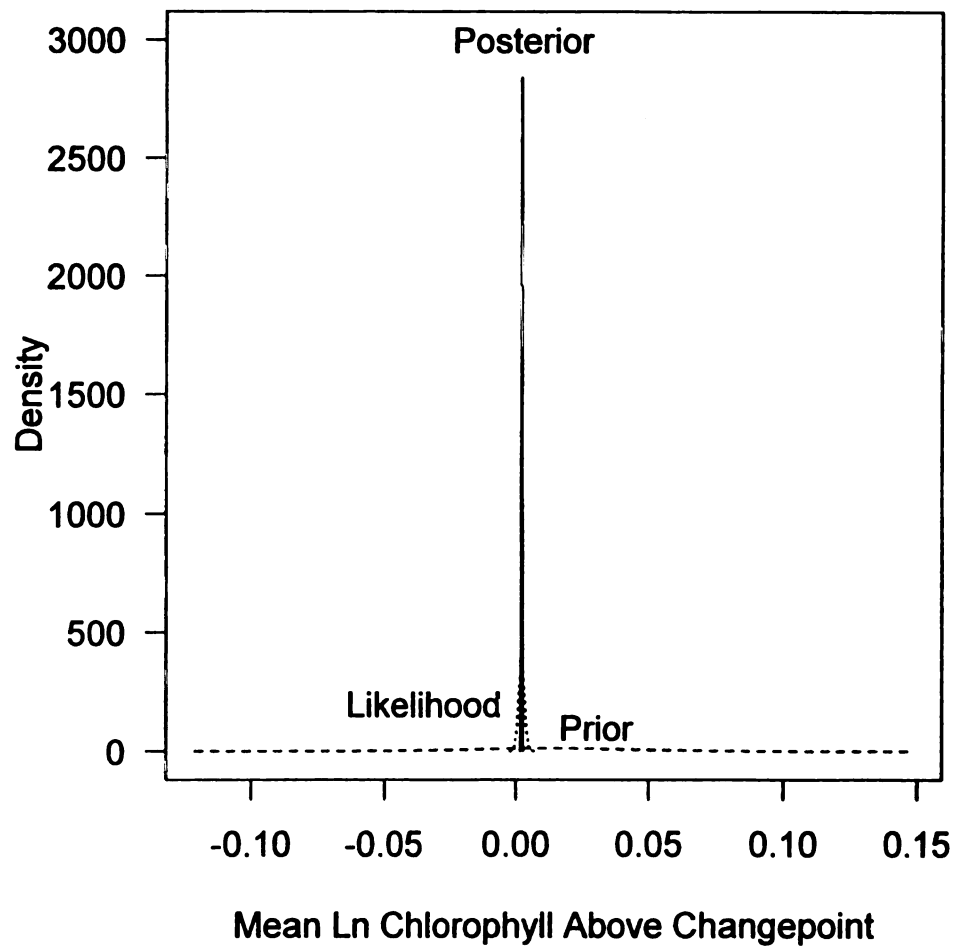


Figure 18. Prior, likelihood, and posterior densities of natural-log-transformed chlorophyll (mg/L) above the total phosphorus threshold.

priors and the data. Uncertainty in the priors and the likelihoods for both of these parameters were similar, so the density of the posterior distributions was located roughly half way between the means of the prior and likelihood distributions (Figures 19 and 20). Prior estimates were lower than likelihoods for each of the land use parameters, so posterior estimates were higher than those suggested by prior information. Priors for channel length, precipitation, and the constant were uninformative, centered on zero and with exceptionally wide distributions; therefore, the posteriors for these parameters were dominated by the data. Land use parameters were positively related to TP concentrations, while channel length and precipitation were negatively related to TP (Table 8).

At the Cass River site, the mean posterior inferred reference TP was 0.020 mg/L with a standard deviation of 0.00789. Inferred reference TP was based on 0.5% land use for both urban and crop gradients because this level was sufficient to prevent the probability of non-attainment from becoming significantly higher than 10% (Figure 14). The probability that the inferred reference TP was greater than the stressor-response threshold based on their posterior distributions was 0.0116. Therefore, the benchmark TP concentration based on $RR=1$ is the TP concentration that represents the 1.16% quantile of the

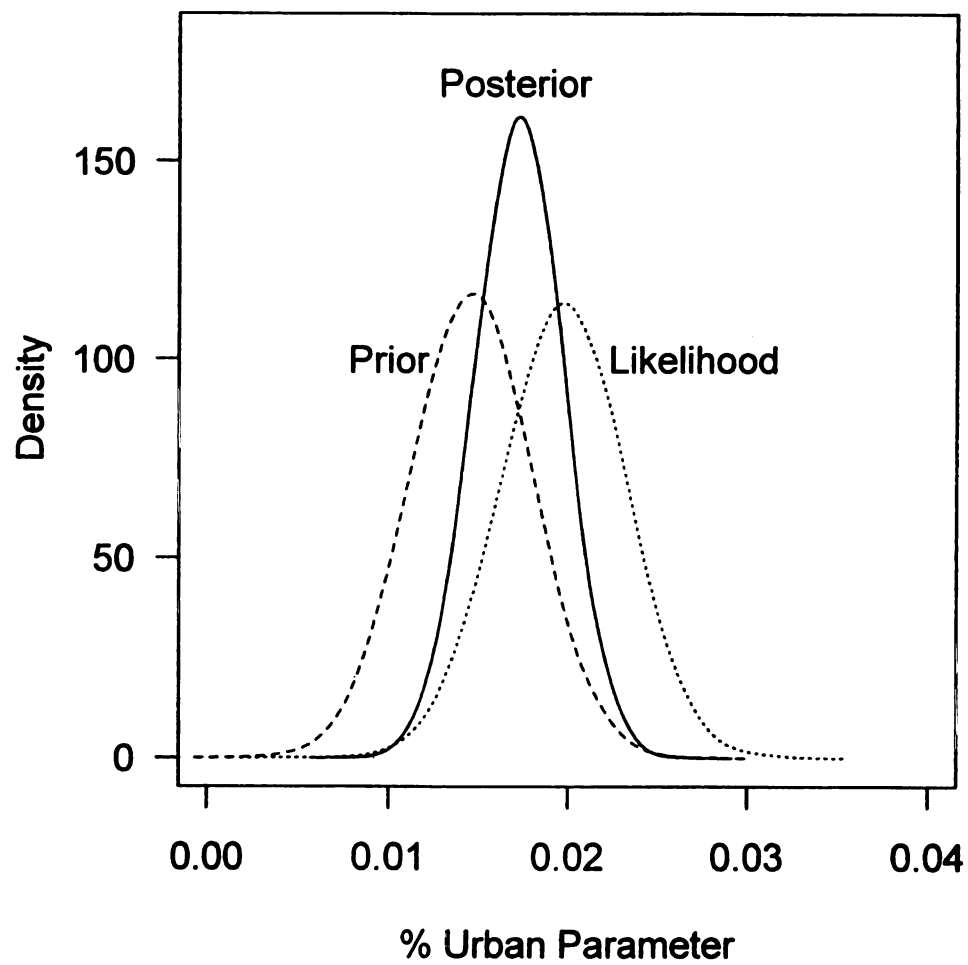


Figure 19. Prior, likelihood, and posterior densities of the predictive model parameter for percent urban land use in the watershed.

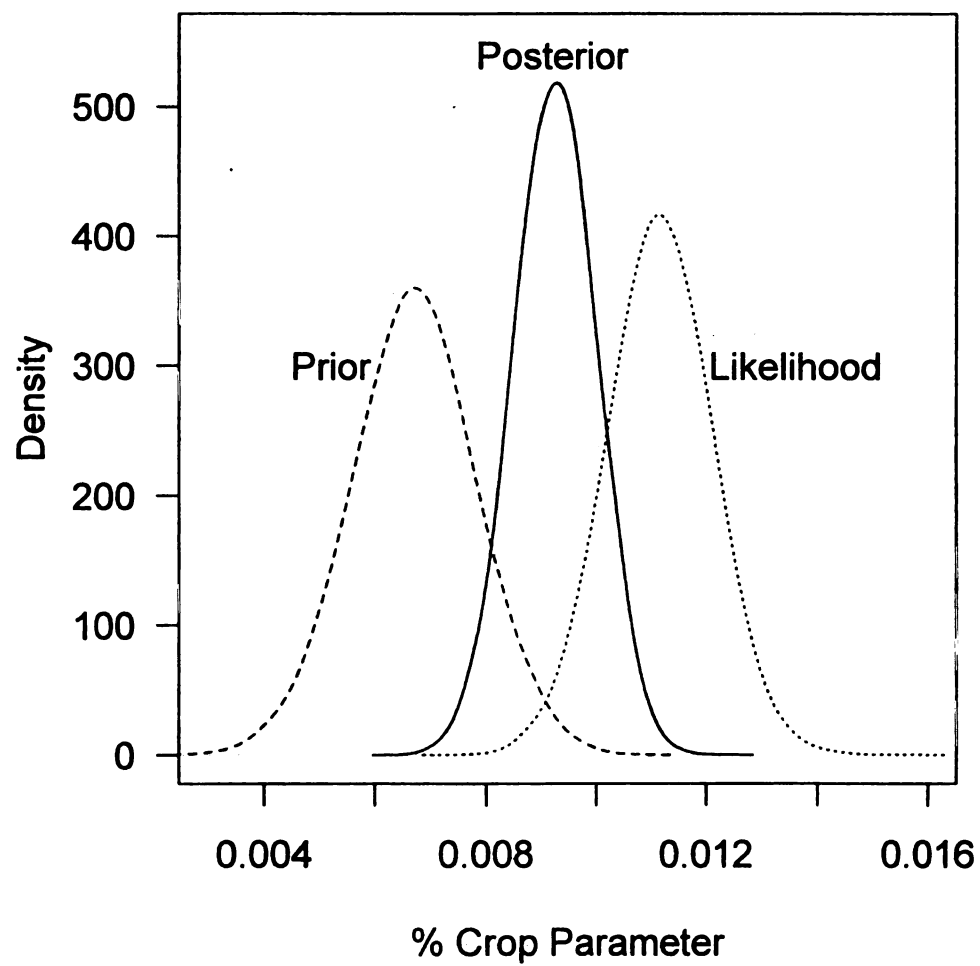


Figure 20. Prior, likelihood, and posterior densities of the predictive model parameter for percent crop land use in the watershed.

Table 8. Posterior estimation of parameters in the predictive model used to infer log-transformed total phosphorus (mg/L), $\log(\text{TP}) = \beta_0 + \beta_1[\ln(\text{Channel Length})] + \beta_2[\ln(\text{Precipitation})] + \beta_3[\% \text{ Urban}] + \beta_4[\% \text{ Crop}] + \varepsilon$.

Parameter	Posterior Percentile						
	2.5%	5.0%	25.0%	50.0%	75.0%	95.0%	97.5%
Constant	3.7709	4.4721	6.8998	8.5297	10.1624	12.5553	13.3273
Ln Channel Length	-0.1170	-0.1076	-0.0789	-0.0594	-0.0395	-0.0108	-0.0014
Ln Precipitation	-2.1910	-2.0770	-1.7201	-1.4762	-1.2340	-0.8764	-0.7620
Percent Urban	0.0126	0.0133	0.0156	0.0172	0.0189	0.0212	0.0219
Percent Crop	0.0078	0.0081	0.0088	0.0092	0.0097	0.0104	0.0107

threshold distribution. This resulted in a benchmark of 0.034 mg/L TP (Figure 21). The estimated risk of exceeding the threshold at the current concentration of 0.090 mg/L TP was 1.0000, 86 times higher than at inferred reference conditions.

Current TP concentrations in the Michigan data average 0.059 mg/L TP. The mean inferred reference TP using mean parameter estimates (i.e., discounting uncertainty) was 0.015 mg/L TP. Roughly 70% of sites were above their predicted TP using this estimation approach, with a mean difference between current levels and mean inferred levels of 0.045 mg/L; most sites, however, are within 0.020 mg/L of predicted reference levels of TP. Approximately 60% of sites are currently below the median posterior threshold value from the stressor-response relationship, whereas all mean posterior inferred TP concentrations were below this value.

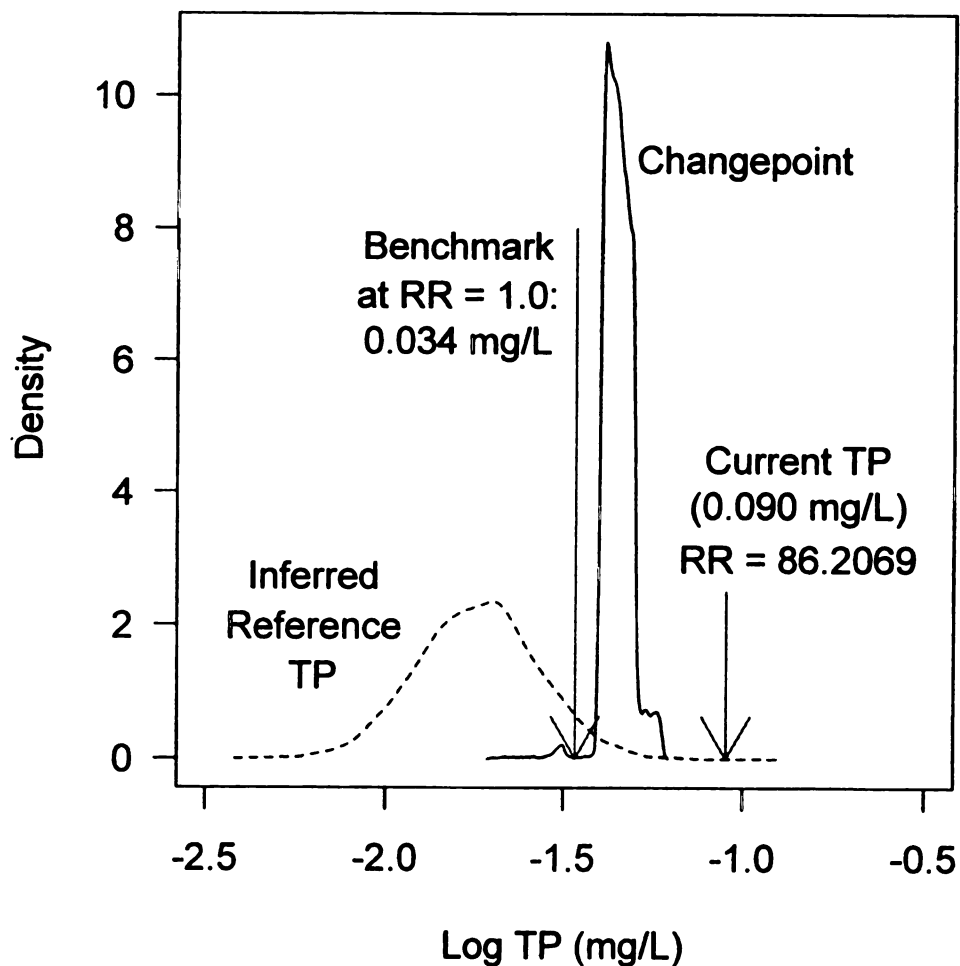


Figure 21. Detailed presentation of the Cass River site analysis, showing posterior densities in predicted reference total phosphorus and the changepoint separating sites characterized by low mean chlorophyll from sites characterized by high mean chlorophyll. Current total phosphorus and the associated relative risk (RR) at that level is shown, along with the TP benchmark determined at a relative risk of one.

DISCUSSION

One of the difficulties in developing nutrient criteria is making sense of different types of information. Two alternative methods for developing candidate nutrient criteria are reference-based approaches and effects-based approaches (i.e., stressor-response). The reference-based approach defines expected conditions by inferring what a site would look like with minimal human disturbance, while the effects-based approach defines expected conditions below some nutrient level that is likely to cause an undesirable change in a valued attribute of the ecosystem. Published research can also support nutrient criteria development, better informing but further complicating the decision making process. I have presented a framework for combining these sources of information to assist in the development of candidate nutrient criteria.

Both reference-based and effects-based methods can result in criteria that are over- or under-protective, and provide an incomplete picture when used alone. The widely applied definition of a reference site is one with minimal human disturbance, characteristic of pre-European conditions. The reference approach applied in this way may be over-protective of societal goals for water quality. On the other hand, if the best available sites in a region are

used to define reference conditions, criteria may be under-protective if similar streams in the region are only represented by disturbed systems. These are the common critiques of the reference-based approach.

Effects-based approaches are presented as an alternative for establishing benchmarks. This method is sensitive to the endpoints selected, which are restricted to available data. Effects based criteria can also be over-protective if valued ecological attributes for which a criterion is intended to protect is not supported under minimally disturbed conditions. Therefore, inferring reference conditions is recommended, even when criteria are set using benchmarks from effects-based approaches. The relative risk framework provides an easy-to-interpret summary of this information.

In this analysis I examined the response of water column chlorophyll a (mg/L) to TP. Chlorophyll was chosen as an endpoint because these data were available for many sites in the National Nutrient Database for the State of Michigan, there is a known causal relationship with TP, and chlorophyll has direct potential aesthetic impacts on stream use. In application, many endpoints that are representative of designated uses should be explored. Additional endpoints may include changes in algal species composition, presence of filamentous benthic algae, or the

presence of nuisance plants. Indirect effects of nutrient enrichment such as decreases in dissolved oxygen, loss of invertebrate taxa, changes in invertebrate metrics, loss of fish, or decreased perception of usability from public surveys may also be used as endpoints. Similarly, nitrogen should also be examined as a potential cause of cultural eutrophication. Dodds and Welch (2000) note that a survey of 158 bioassays found 13% showed stimulation by N alone, 18% by P alone, 44% by both N and P, and 25% by neither nutrient.

The TP-chlorophyll stressor-response relationship was empirically derived, rather than experimental. This approach has benefits and drawbacks. Causality can not be established definitively in the observed relationship, however, increased algal production resulting from increased phosphorus in experimental settings has been observed in several studies (Rier and Stevenson, submitted; Hillebrand 2002; Bothwell 1988, 1989). King and Richardson (2003) present an approach for developing nutrient water-quality criteria that uses mesocosm experiments to support thresholds found in observational relationships. This is a useful approach if mesocosms provide a relatively realistic representation of the natural system. Within the framework presented here, these data could be used as prior information in the Bayesian threshold analysis.

Experiments examining phosphorus-algal biomass relationships in experimental stream settings at several levels of phosphorus are rare. Preliminary evaluation of experimental results that were available (Rier and Stevenson, submitted; Bothwell 1988, 1989) suggested that they were inappropriate for characterizing threshold in natural settings. In all cases, the most obvious threshold took place between the control and the lowest phosphorus treatment, which ranged from 0.0001-0.002 mg P/L. This is well below the threshold likely to exist in natural systems due to human disturbance, as natural levels of TP are probably several times greater.

Although general patterns for total phosphorus in Michigan streams relative to inferred reference conditions are summarized for the full dataset in the results section, these data should be interpreted with caution because they do not adequately account for uncertainty. Although TP at many sites is above the mean predicted reference concentration for the site, many sites are probably within the 95% prediction interval for the site. Uncertainty involved in inference of reference TP and the TP threshold warrant examination of sites individually, rather than as a population. The Cass River analysis provides an example for this approach.

Well-informed environmental management decisions should

make use of relevant and available information, including the uncertainty in predictive models. The Bayesian approach presented here uses prior information and attempts to quantify uncertainty in the inference of reference conditions and ecological thresholds. Uncertainty may still be underestimated here, in part because within-site sources of variation are not accounted for in this approach. These sources of variation include temporal fluctuations in TP and chlorophyll, uncertainty associated with the precision of analytical methods for measuring TP and chlorophyll, and with the accuracy of the measurements related to sample handling. Hierarchical Bayesian methods can help address sources of variation occurring within sampling units (Clark 2003) and could improve the approach shown here. Additionally, the land use-attainment relationships for defining reference conditions underestimate uncertainty. Category 2 attainment in Michigan's 305(b) list means that a site is attaining for all attributes tested. Ideally, category 1 attainment would be used. Category 1 attainment means that a site is attaining all designated uses. No sites have been evaluated at this level in Michigan. The approach presented here does, however, make significant strides toward quantifying uncertainty and provides a formal method for integrating reference-base and effects-based approaches for water

quality criteria development. The relative risk approach provides a useful way to communicate site-specific results, while incorporating various sources of uncertainty in a complex set of models.

REFERENCES

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 268-281 in Petrov, B. N., and Csaki, F., eds. 2nd International Symposium on Information Theory.
- Akaike, H. 1985. Prediction and entropy. Pages 1-24 in Atkinson, A. C., and Fienberg, S. E., eds. A Celebration of Statistics. Springer Verlag, New York, NY.
- Bolstad, W. M. 2004. Introduction to Bayesian Statistics. John Wiley and Sons, Inc., Hoboken, NJ.
- Bothwell, M. L. 1988. Growth rate responses of lotic periphytic diatoms to experimental enrichment: the influence of temperature and light. Canadian Journal of Fisheries and Aquatic Sciences 45: 261-270.
- Bothwell, M. L. 1989. Phosphorus-limited growth dynamics of lotic periphytic diatom communities: areal biomass and cellular growth rate responses. Canadian Journal of Fisheries and Aquatic Sciences 46: 1293-1301.
- Brewer, C. A., and Gross, L. J. 2003. Training ecologists to think with uncertainty in mind. Ecology 84: 1412-1414.
- Cairns, J., Jr. 1983. The case for simultaneous toxicity testing at different levels of biological organization. Pages 111-127 in Bishop, W. E., Cardwell, R. D., and Heidolph, B. B., eds. Aquatic Toxicology and Hazard Assessment: Sixth Symposium. American Society for Testing Materials, Philadelphia, PA.
- Cairns, J., Jr., and Pratt, J. R. 1986. Use of protozoan communities in protecting aquatic ecosystems. Symposia Biologica Hungarica 33: 187-197.
- Clark, J. S. 2003. Uncertainty in ecological inference and forecasting. Ecology 84: 1349-1350.
- Carpenter, S. R. 2002. Ecological futures: building an ecology for the long now. Ecology 83: 2069-2083.

- Dodds, W. K., and Oakes, R. M. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. *Limnology and Oceanography Methods* 2: 333-341.
- Dodds, W. K., Smith, V. H., and Lohman, K. 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., and Welch, E. B. 2000. Establishing nutrient criteria in streams. *Journal of the North American Benthological Society* 19: 186-196.
- Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications* 6: 1036-1046.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1995. *Bayesian Data Analysis*. Chapman and Hall, Boca Raton, FL.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. 1996. Introducing Markov chain Monte Carlo. Pages 1-19 in Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., eds. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, Boca Raton, FL.
- Gotelli, N. J., and Ellison, A. M. 2004. *A Primer of Ecological Statistics*. Sinauer Associates, Inc., Sunderland, MA.
- Hastie, T. J., and Tibshirani, R. J. 1990. *Generalized Additive Models*. Chapman and Hall, New York, NY.
- Hillebrand, H. 2002. Top-down versus bottom-up control of autotrophic biomass—a meta-analysis on experiments with periphyton. *Journal of the North American Benthological Society* 21: 349-369.
- King, R. S., and Richardson, C. J. 2003. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management* 31: 795-809.
- Landis, W. G., Duncan, P. B., Hayes, E. H., Markiewicz, A. J., and Thomas, J. F. 2004. A regional retrospective assessment of the potential stressors causing the decline of the Cherry Point Pacific herring run. *Human and Ecological Risk Assessment* 10: 271-297.

- Peterson, G. D., Carpenter, S. R., and Brock, W. A. 2003. Uncertainty and the management of multistate ecosystems: an apparently rational route to collapse. *Ecology* 84: 1403-1411.
- Pielke, R. A., Jr., and Conant, R. T. 2003. Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology* 84: 1351-1358.
- Qian, S. S., King, R. S., Richardson, C. J. 2003. Two statistical methods for the detection of environmental thresholds. *Ecological Modelling* 166: 87-97.
- Reckhow, K. H. 1990. Bayesian inference in non-replicated ecological studies. *Ecology* 71: 2053-2059.
- Reckhow, K. H., Arhonditsis, G. B., Kenney, M. A., Hauser, L., Tribo, J., Wu, C., Elcock, K. J., Steinberg, L. J., Stow, C. A., and McBride, S. J. 2005. A predictive approach to nutrient criteria. *Environmental Science and Technology* 39: 2913-2919.
- Rier, S. T., and Stevenson, R. J. Submitted. Response of periphytic algae to gradients in nitrogen and phosphorus in streamside mesocosms.
- Smith, R. A., Alexander, R. B., and Schwarz, G. E. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. *Environmental Science and Technology* 37: 2039-3047.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. 2003. BUGS: Bayesian inference Using Gibbs Sampling, Version 1.4.1. Cambridge: Medical Research Council Biostatistics Unit.
- Stevenson, R. J., Bailey, R. C., Harrass, M. C., Hawkins, C. P., Alba-Tercedor, J., Couch, C., Dyer, S., Fulk, F. A., Harrington, J. A., Hunsaker, C. T., and Johnson, R. K. 2004. Interpreting results of ecological assessments. Pages 85-111 in Barbour, M. T., Norton, S. B., Preston, H. R., and Thornton, K. W., eds. *Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making*. Society of Environmental Toxicology and Chemistry.
- United States Environmental Protection Agency (USEPA). 1996. National Water Quality Inventory 1996 Report to Congress. Office of Water, U.S. Environmental Protection Agency. EPA 841-R-97-008.

United States Environmental Protection Agency (USEPA).
1998. Clean Water Action Plan: Restoring and
Protecting America's Waters. U.S. Environmental
Protection Agency. EPA-840-R-98-001.

United States Environmental Protection Agency (USEPA).
2000. Nutrient Criteria Technical Guidance Manual:
Rivers and Streams. Office of Water, Office of Science
and Technology, U.S. Environmental Protection Agency.
EPA-822-B-00-002.

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

Wade, P. R. 2000. Bayesian methods in conservation biology.
Conservation Biology 14: 1308-1316.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02736 7725