Data Analysis to Support Development of Nutrient Criteria for Maryland Free-Flowing Waters

Final Report

Submitted 15 July 2011
Edits and clarifications incorporated September 6 and October 20, 2011, **and June 1, 2012**
Table18 and 29 corrected November 3, 2011

Ross Mandel
Claire Buchanan
Adam Griggs
Andrea Nagel
Olivia Devereux

Prepared for

Maryland Department of the Environment
Science Services Administration
1800 Washington Blvd
Baltimore, MD 21230

by

Interstate Commission on the Potomac River Basin
51 Monroe St, PE-08
Rockville, MD 20850
301-984-1908
www.potomacriver.org

ICPRB Report 11-2

To receive additional copies of this report, please write or call

Interstate Commission on the Potomac River Basin
51 Monroe St., PE-08
Rockville, MD 20182
301-984-1908

After the report has been reviewed and released for distribution by the Maryland Department of the Environment, an electronic copy of the report can be downloaded from the Publications tab at www.potomacriver.org.

Disclaimer

Acknowledgements

This project was supported by a grant from the Maryland Department of the Environment (MOU U00P0400030) and by the Interstate Commission on the Potomac River Basin, an interstate compact river basin commission of the United States government and the compact's signatories: Maryland, Virginia, Pennsylvania, West Virginia, and the District of Columbia. The source of the Maryland Department of the Environment funding was federal monies provided to the state through the American Recovery and Reinvestment Act (ARRA) of 2009.

The primary ("raw") data sets used in the analysis were provided by the state agencies of Pennsylvania, West Virginia, Maryland, Delaware, and Virginia. Additional data and data products, including modeling output, were provided by the Chesapeake Bay Program Office. The authors wish to acknowledge and thank ICPRB staff Jan Ducnuigeen, Sunghee Kim, Jacqueline Johnson, and Michael Mallonee for their GIS and data management contributions to the project, and MDE staff Adam Rettig for his guidance, project participation, and assistance in acquiring data sets. MDE Science Services Administration feedback to ICPRB during periodic meetings to report progress was especially useful and informative.

Finally, we wish to recognize the monitoring program "crews" without whom this study would not have been possible. Their dedicated hard work makes science-based planning and resource management possible.

Executive Summary

The Interstate Commission on the Potomac River Basin (ICPRB) performed several data analyses to assist Maryland Department of the Environment (MDE) in its efforts to develop total phosphorus and total nitrogen criteria for the state's non-tidal free-flowing streams and rivers. A stressor-response approach was used and three different segments of the aquatic community were targeted: phytoplankton, periphyton, and macroinvertebrates. A diverse data set was assembled for each aquatic community from available data in the Mid-Atlantic region, which were primarily state monitoring program data. The analysis tried to address two of the most significant criticisms of the stressor-response approach, namely 1) nutrient concentration thresholds indicative of impairment cannot be confidently identified because confounding factors introduce so much variability in relationships between stressor and response variables, and 2) it is difficult to relate nutrient concentrations breakpoints or thresholds directly to aquatic life use support.

Some natural variability was reduced by classifying biological response metrics by physiographic region and stream size class. A recursive partitioning (RPART) analysis technique was then employed to identify confounding factors. RPART is a non-parametric decision tree approach that splits the data based on which independent (stressor) variables optimally differentiate observations in the dependent (response) variable. Breakpoints are identified for the independent variables most capable of splitting the response variable data into increasingly homogeneous groups. These breakpoints were used to identify data records for which the response variable either co-varies with a non-nutrient factor or is confounded by non-nutrient factors. Nutrient responses of phytoplankton were not evident unless the water clarity surrogates turbidity and dissolved organic carbon were simultaneously considered. Periphyton nutrient responses in 1^{st} – 4^{th} order streams were evident only after records associated with marginal/poor stream bank conditions and high conductivity were removed. Macroinvertebrate nutrient responses in 1^{st} – 4^{th} order streams were evident only after records associated with high conductivity and marginal/poor in-stream habitat quality, and to a less degree extreme pH levels and low dissolved oxygen, were removed. Once the confounded data records were removed from the analysis data set, response variables in the remaining data from "minimally disturbed locations" demonstrated clear nutrient thresholds. A "binning approach" was then used to group the data into distinct nutrient categories, allowing the biological community's nutrient responses to be examined in the context of relatively undisturbed, naturally varying environmental conditions.

Nutrient thresholds protective of high quality biological communities were 0.012 – 0.087 mg/liter for water column TP and 0.58 – 2.67 mg/liter for water column TN. Thresholds varied by physiographic region, stream size, and—in the case of phytoplankton—water clarity. These ranges of protective thresholds agree with thresholds identified by other researchers.

There are no impairment criteria for phytoplankton or periphyton in Maryland regulations, hence nutrient thresholds for impairment of a designated aquatic life use could not be determined for these two biological groups. Maryland uses a benthic macroinvertebrate index of biotic integrity (BIBI) as a biocriterion for determining impairment of aquatic life uses in 1^{st} - 4^{th} order streams. Scientifically defensible nutrient thresholds of macroinvertebrate impairment were difficult to identify because they could not be untangled from the impacts of other stressors. However, a broad variety of macroinvertebrate metrics are sensitive to nutrient concentrations once confounding factors are accounted for or removed, so it is likely nutrient impacts are acting in concert with other stressors in heavily degraded streams and rivers.

The results of this study can best be interpreted in the context of the Tiered Aquatic Life Uses (TALUs) concept and the associated Biological Condition Gradient (BCG) concept. Distinct differences in biological condition—the probability of algal blooms and macroinvertebrate impairment—are associated with sometimes overlapping ranges of multiple physical and chemical parameters, including nutrients. The ranges of environmental condition collectively describe at least three categories of stress. If connections between specific levels of macroinvertebrate and periphyton status, and possibly between macroinvertebrate and phytoplankton status, can be established using monitoring program parameters, periphyton and phytoplankton indicators could be incorporated into the Maryland biological stressor identification process as new measures of nutrient degradation.

The following monitoring and analysis recommendations came from the study results:

- Obtain Washington Aqueduct algal monitoring data and analyze relation between chlorophyll *a* concentrations and algal taxa;
- Obtain ORSANCO algal monitoring data and analyze relation between chlorophyll *a* concentrations and algal taxa;
- Using the methods of this study, analyze relation between benthic metrics as the dependent variable and periphyton chlorophyll *a* content, phosphorus content, or ash-free dry mass as independent variables using VADEQ and SRBC periphyton monitoring data;
- Try to associate chlorophyll *a* monitoring data in the Coastal Plain with macroinvertebrate monitoring data and, if successful, determine relation between macroinvertebrate metrics and observed chlorophyll *a* concentrations;
- Explore relation between MBSS reference conditions and "minimally disturbed location" (MDL) samples used in this study;
- Test sensitivity of MBSS fish index of biotic integrity (FIBI) across the study's nutrient bins;
- Develop a Nutrient Biotic Index using nutrient-sensitive benthic metrics, chlorophyll *a* water column concentrations, and or chlorophyll *a* periphyton concentrations; and
- Explore potential to develop BCG tiers and a stressor gradient which incorporates the habitat, water quality, and biological thresholds identified in this study.

Implications to non-tidal stream and river nutrient concentrations of the Chesapeake Bay Total Maximum Daily Load (TMDL) nutrient reduction allocations are discussed. Currently, delays in the Chesapeake Bay Program Phase 5 Watershed Model runs prevent testing of specific TMDL reduction scenarios. An example of the effects on non-tidal streams and rivers of implementing the nutrient reductions was performed on the Upper Pocomoke River 8-digit watershed.

Table of Content

Abbreviations

ALTER	Channel alteration
ARRA	American Recovery and Reinvestment Act of 2009
AFDM	Ash-free dry mass (mg/m^2)
BANKS	Bank stability
BANKVEG	Vegetative protection on stream bank or near-stream portion of riparian zone
BCG	Biological Condition Gradient
BIBI	Biological (or Benthic) Index of Biotic Integrity
BSID	Biological Stressor Identification
CART	Category and Regression Tree
CHL_BEN	Periphyton chlorophyll *a* content (mg/m^2)
COND	Specific conductivity
CWA	United States Clean Water Act
DE DNREC	Delaware Department of Natural Resources and Environmental Control
DIN	Dissolved inorganic nitrogen
DO	Dissolved oxygen (mg/liter)
DOC	Dissolved organic carbon (mg/liter)
GCG	Generalized Stressor Gradient
HQI	Habitat Quality Index
IBI	Index of Biotic Integrity
ICPRB	Interstate Commission on the Potomac River Basin
IQR	Inter-quartile (25^{th} percentile - 75^{th} percentile) range
MACP	Mid-Atlantic Coastal Plain
MAL	Minimum Allowable Limit
MBSS	Maryland Biological Stream Survey
MDE	Maryland Department of the Environment
MDL	Minimally disturbed location
MDDNR	Maryland Department of Natural Resources
MOU	Memorandum of Understanding
NEAHCS	Northeast Aquatic Habitat Classification System
NTU	Nephelometric Turbidity Units
ORSANCO	Ohio River Valley Water Sanitation Commission
PADEP	Pennsylvania Department of Environmental Protection
PRV	Combined Piedmont, Ridges, and Valleys bioregions
RBP	Rapid Bioassessment Protocols
RIPVEG	Riparian zone width with natural vegetation
RPART	Recursive partitioning
SEP	Southeastern Plain
SRBC	Susquehanna River Basin Commission
TALU	Tiered Aquatic Life Use
TMDL	Total Maximum Daily Load
TN	Total nitrogen (mg/liter)
TP	Total phosphorus (mg/liter)
TP_ALG	Periphyton total phosphorus content (mg/liter)
TSS	Total suspended solids (mg/liter)
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
VADEQ	Virginia Department of Environmental Quality
WIP	Watershed Implementation Plan
WVDEP	West Virginia Department of Environmental Protection

1. Introduction

1.1 General Background

Under the United States Clean Water Act (CWA), water quality standards are instituted to protect the designated uses of the nation's lake, reservoirs, estuaries, rivers and streams. Designated uses include primary contact recreation and protection of aquatic life—the "fishable and swimmable" goals of the CWA—but also include such uses as drinking water supply and shellfish propagation and harvest. A water quality standard consists of a designated use and narrative or numeric criteria specifying the conditions necessary to protect that use. A narrative criterion can be as general as the specification that no material may pollute the waters of a State in amounts sufficient to create a nuisance or interfere with designated uses. In contrast, numerical criteria set quantitative limits on the concentration of materials. An example is the criterion for dissolved oxygen (DO) for warm-water fisheries, which states that DO concentrations must be above 5 mg/l at all times.

The 1996 National Water Quality Inventory reported that nutrients were the second most significant cause of the impairment of rivers and streams, contributing to 40% of the reported impairments. To address the nutrient impairment of rivers and streams the U. S. Environmental Protection Agency's (USEPA) 1998 Clean Water Action Plan called for the development of numeric criteria for nutrients by states, tribes, and territories. USEPA envisions potential numerical nutrient criteria not only for the nutrients nitrogen and phosphorus but also for nutrient response variables such as algal biomass, chlorophyll *a*, and Secchi depth.

According to Grumbles (2007) the implementation of numerical nutrient criteria will have the following benefits:

- Easier and faster development of TMDLs;
- Quantitative targets to support trading programs;
- Easier to write protective NPDES permits;
- Increased effectiveness in evaluating success of nutrient runoff minimization programs; and
- Measureable objective water quality baselines against which to measure environmental progress.

To facilitate development of numerical nutrient criteria, USEPA has issued a series of guidance documents and recommendations for nutrient criteria by ecoregion. USEPA recognized several different methods by which nutrient criteria could be established, including (1) using summary statistics from reference reaches to set nutrient criteria; (2) setting nutrient criteria on the basis of published scientific studies; and (3) establishing criteria on the basis of predictive relationships between nutrients and the aquatic biological community through data analysis, also known as the stressor-response approach.

The stressor-response approach analyzes existing water quality data to determine relationships between nutrients as independent stressor variables and measures of water quality as dependent response variables. These "response variables" could include measures of algae such as chlorophyll *a* concentration, cell counts, or biomass, or benthic macroinvertebrate metrics used in water quality assessment such number of taxa, percent dominant taxa, the Shannon Wiener diversity Index, etc. The relationship could be demonstrated through classical statistical analysis such as linear regression, or with newer forms of statistical analysis such as quantile regression, ordination methods, and canonical correlation analysis. USEPA (2010c) provides guidance on how the stressor-response approach can be used to develop scientifically-defensible nutrient criteria.

1.2 Project Overview

This report documents a data analysis done by the Interstate Commission on the Potomac River Basin (ICPRB) to develop and support recommendations for nutrient criteria for Maryland's non-tidal free-flowing streams and rivers using the stressor-response approach. The nutrients in question are phosphorus and nitrogen—two essential elements in the food web which can pollute aquatic ecosystems if they are present in concentrations that cannot be rapidly incorporated into and distributed throughout the food web. Three different segments of the aquatic community were targeted for analysis: (1) phytoplankton, (2) periphyton, and (3) benthic macroinvertebrates. In accordance with the River Continuum Concept (Vannote *et al*. 1980) which explains how longitudinal physical changes in streams and rivers govern the structure and activities of aquatic communities, phytoplankton (free-floating algae) tend to be the dominant in-stream primary producer in large rivers where the riparian canopy does not limit the availability of light and the water column is deep enough to create pelagic conditions. Periphyton (attached algae), along with rooted aquatic plants, tend to be the dominant in-stream primary producers in mid-sized streams and small rivers where the riparian canopy partially shades the width and depth is shallow enough to allow light sufficient for photosynthesis to penetrate to the bottom. Since the impairment of aquatic communities by nutrients is often through excess primary production, the most direct impact of excess nutrients should be among in-stream primary producers. It should be evident as an increase in algae biomass and chlorophyll *a* concentrations and a greater magnitude of diel change in pH and dissolved oxygen.

Bioassessment of benthic macroinvertebrates has been established as a method of assessing the overall health of aquatic communities, and metrics and indices based on macroinvertebrate samples have long been integrated into formal procedures for determining the water quality status of rivers and streams, as required by the Clean Water Act. Although the impacts of excess nutrients on a macroinvertebrate community are less direct than their impacts on primary producers, macroinvertebrates provide a measure of the health of the animal community and a more direct means of assessing whether excess nutrients are undermining the ability of a river or stream to support aquatic life, as the Clean Water Act mandates.

A diverse dataset was assembled for each aquatic community targeted for analysis. For phytoplankton the primary dataset consisted of over 7,000 records with concurrent chlorophyll *a* (Chla), total nitrogen (TN), total phosphorus (TP), and turbidity observations. The data were collected by the Maryland Department of Natural Resources (MDDNR) Core Trend monitoring program, by MDE in special studies, and by the Delaware Department of Natural Resources and Environmental Control (DE DNREC) biological monitoring program. The primary dataset used to analyze relationships between nutrients and benthic macroinvertebrates was a database of macroinvertebrate, habitat, and water quality monitoring results assembled to support the development of the Chesapeake Bay Program's benthic index of biological integrity ("Chessie BIBI"). The Chessie BIBI database contains information collected by 23 local, state, and federal agencies across the entire Chesapeake Bay basin. A subset of approximately 8,000 sampling events from the physiographic provinces found in Maryland was used in the nutrient criteria analysis. MDE and ICPRB staff assembled existing periphyton datasets collected by MDDNR, DE DNREC, Virginia Department of Environmental Quality (VADEQ), Pennsylvania Department of Environmental Protection (PADEP), the Susquehanna River Basin Commission (SRBC), and MDE with the U.S. Geological Survey (USGS). Periphyton sampling is not performed routinely by government agencies, and systematic analysis of these data was hampered by differences among the programs in the parameters measured to characterize periphyton and in the constituents measured concurrently with the periphyton collection. The analysis concentrated on the VADEQ dataset, which had over 100 observations of periphyton biomass and chlorophyll *a* which could be linked to measurements of habitat, benthic macroinvertebrates, water column nutrients and other constituents.

The stressor-response approach to developing nutrient criteria has been subject to much criticism. Two of the most significant criticisms of the approach are that (1) nutrient concentration thresholds indicative of impairment cannot be confidently identified because confounding factors introduce so much variability in relationships between the stressor and response variables; and (2) it is difficult to relate nutrient concentration breakpoints or thresholds directly to aquatic life use support. The basis for the first criticism is illustrated in **Figure 1**, which shows the response of Beck's Index, a macroinvertebrate index of pollution-sensitive taxa, to a range of TP concentrations. High TP concentrations are associated with low index values but low TP concentrations are associated with a wide range of index values. This wedge-shaped response is typical of many biological variables when plotted against nutrient gradients. Regressions can be statistically significant but of little use in explaining the wedge's variance at the low end of the nutrient gradient. The narrow range of Beck's Index values at the high TP end can be due to direct stress by the nutrient or may be evidence that TP is positively correlated with some other stressor or anthropogenic disturbance. When TP concentrations decrease, stress caused by the nutrient or the associated factor is relieved and other environmental factors assume control of the biological community. If some of those factors are additional stressors that behave independently of TP, a wide range of index scores is observed and a clear threshold for the nutrient stressor cannot be found. It is therefore incumbent on an implementation of the stressor-response approach to take into account confounding variables. Confounding variables can be other stressors such as poor physical habitat or low pH. They can also be natural factors that add variability to the relation between nutrient stressor and biological response. Variability by natural factors can be minimized by subdividing the data into more homogenous groups. The application of ecoregion, for example, is intended to capture the effects of natural geographic variability on biological communities that are independent of human influence. Similarly, according to the River Continuum Concept, biological communities vary naturally with river size and should be analyzed in that context.

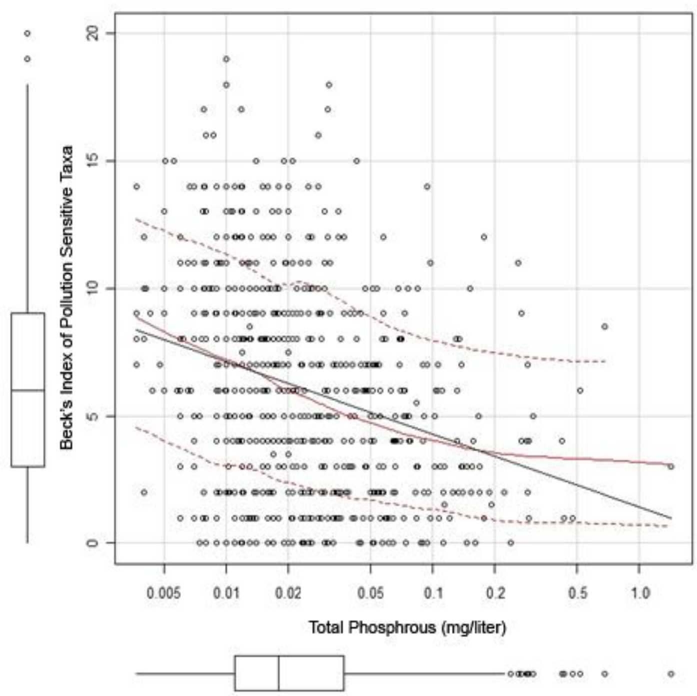


Figure 1. Scatter plot of Beck's Index of pollution-sensitive family-level taxa (normalized to a 100-count sample) *versus* total phosphorus concentration in the Piedmont physiographic region. Data have not been filtered for confounding factors. Black solid line, linear regression; red solid line, LOWESS curve; red dashed lines, 90% confidence interval around the LOWESS curve.

Once the effects of important confounding factors have been recognized and accounted for and a biological response metric has been related to a nutrient gradient, there remains the distinct step of using the relationships to identify nutrient concentrations likely to impair aquatic life. This can be difficult if significant unexplained variability remains in the response metric after the known confounding variables are taken into account. Combining several response metrics into a composite index can reduce the unexplained variability. For example, jackknife validations of the Chessie BIBI index produce much smaller total errors than the individual metrics comprising the index, indicating that error inherent in one metric's ability to identify high quality sites is outweighed by correct identifications made by the index's other metrics (Buchanan *et al.* 2011). Another difficulty is deciding when impairment of aquatic life occurs. Biological communities can tolerate a certain amount of intermittent stress and recover with no lasting ill effects. Impairment is the result of repeated and prolonged stress that changes biological community structure and function in recognizable ways. *Degrees* of change rather than an abrupt change are typically observed in biological responses to stressors, so the decision as to what constitutes nutrient impairment of a biological community is often a statistical or political one.

The methodology adopted for this project tries to address the first of the two major criticisms of the stressor-response approach. An initial exploratory analysis identified or confirmed the environmental factors that exert strong controls on the response variables. These are most likely to confound the nutrient stressor-response relationships. Analysis techniques used in the exploratory analysis ranged from scatter plots, longitudinal plots and Spearman Rank correlation tests to recursive partitioning (RPART) routines. (RPART implements in R software many of the Classification and Regression Tree (CART) routines developed for S-plus software and also shows results as binary trees.) Scatter plots, longitudinal plots, and summary statistics were done in Excel 2007 and/or R software. Spearman Rank correlations and RPART tests were performed with the R software. The RPART program implements a non-parametric decision tree technique that uses a collection of user-prescribed rules to split the data based on which independent (stressor) variables optimally differentiate observations in the dependent (response) variable. We used the package RPART and the RAndFriends statistical package version 2.12.2 (Baier and Neuwirth 2007). RPART requires the user to input a simple model with a dependent response variable and two or more independent variables. Independent variables may be categorical, or ratio-scale where numerical breakpoints are selected that divide the data into smaller and more homogeneous nodes. Each split of the data is referred to as a branch, with the final nodes being termed leaves, and the entire set of breaks and nodes forming the tree.

If significant differences in the stream response variables were apparent between physiographic region or stream size, which are categorical features, the analysis datasets were split accordingly and analyzed separately. If the confounding effects of an environmental factor were expressed at levels above or below specific thresholds (e.g. pH > 9 or < 6), records associated with those levels were eliminated from the analysis dataset. If the confounding environmental factor was influential at all nutrient levels (e.g., light energy for photosynthesis), responses to that factor were considered in conjunction with responses to nutrients. This latter approach was successfully employed in the Chesapeake Bay estuary to determine the levels of light (Secchi depth), dissolved inorganic nitrogen (DIN), and ortho-phosphate (PO_4) most likely to support algal blooms (Fisher and Gustafson 2003, Buchanan *et al.* 2005).

After the effects of the most important confounding environmental factors were minimized or accounted for, the RPART program was used to build regression models that recursively partitioned each of the resulting datasets into increasingly homogenous groups. Breakpoints identified by the RPART were then used by the data analyst to create “bins” that have unique environmental conditions and distinct biological responses that grade toward impairment. The RPART breakpoints serve as boundaries of the bins. Each bin is an ensemble of a set of nutrient conditions and the associated levels of one or more confounding factors. The binning approach classifies the overall environmental condition, and the biological response associated with each bin is a response to particular set of nutrient concentrations within the context of the other variables that characterize the bin. The bins thus represent the combination of environmental factors, including nutrients, which best explain the associated biological response.

Nutrient boundaries in the bins that exhibit the greatest biological stress indicate thresholds that can serve as candidate nutrient criteria. It is still necessary, however, to take the explicit step of relating these nutrient thresholds to overall aquatic life use support. In this step, it is important to identify the uncertainty associated with each candidate criterion and the degree of protection of use support associated with that uncertainty.

In Maryland, adopting nutrient criteria for free-flowing rivers and streams is complicated by the fact that the total maximum daily loads (TMDLs) for nutrients to protect Chesapeake Bay call for substantial reductions in both nitrogen and phosphorus loads throughout almost the entire state. The implementation of nutrient reductions under the Bay TMDLs will require a substantial effort on the part of farmers, wastewater treatment plants, and municipal stormwater systems. For any candidate nutrient criteria, it is important to determine whether the non-tidal tributary nutrient reductions required by the Bay TMDLs will be sufficient to protect aquatic life in rivers and streams locally, or whether additional reductions would have to take place in some locations in order to support the aquatic life use in the rivers and streams themselves and not just the Chesapeake Bay.

1.3 Report Overview

The report is divided into eight chapters, including this introduction. Chapter 2 describes the types of data employed in the various analyses. Chapter 3 provides an overview of the conceptual framework for the analyses. It includes a review of the salient features of the River Continuum Concept and how it is expressed in Strahler stream order, as well as the potential influence of ecoregion on both stressor and response variables. Chapter 3 also provides a general discussion on potential confounding factors, such a pH, conductivity, and habitat quality, which can also impact the biota. Chapters 4, 5 and 6 present the analytical results for phytoplankton, periphyton, and macroinvertebrates, respectively. Each chapter describes the results of the exploratory data analysis, RPART analysis, binning approach, and the derivation of candidate nutrient thresholds. Chapter 7 provides recommendations for nutrient criteria that can be supported on the basis of the analyses in Chapters 4, 5 and 6. It also includes recommendations for additional monitoring and analysis to refine recommendations and increase the scientific support for recommended nutrient criteria. Chapter 8 discusses the possible implications of adopting these nutrient criteria on the Chesapeake Bay TMDLs and their associated Watershed Implementation Plans (WIPs), and provides an example of how nutrient criteria might be applied to a HUC-8 watershed.

2. Data types

Three types of data were used as biological response variables in this study: phytoplankton, periphyton and macroinvertebrates. Water quality monitoring programs commonly estimate phytoplankton abundance from chlorophyll *a* (Chla) concentrations measured in surface or depth-integrated water samples. Chla measurements have a long track record in lotic waters of the Chesapeake Bay basin. Phytoplankton taxonomic count data are comparatively sparse and so they were not considered for this study. Macroinvertebrate sample collection also has a long track record in Chesapeake Bay basin lotic waters. All the state agencies perform taxonomic counts to either family- or genus-level and use the results in a variety of indicators and indexes of biotic integrity (IBIs). ICPRB and CBP recently assembled into a common relational database structure ("Chessie" database) macroinvertebrate count data and associated habitat and water quality data submitted by 23 programs. The database was used in this study. Periphyton measurements are relatively new for monitoring programs in the area and the same suite of parameters are not measured by each collecting agencies. Three types of "bulk" measurements are made from material scraped from submerged hard surfaces: Chla content of the material, total phosphorus content of the material, and ash-free dry mass of the material. Some programs also analyze the scraped material for periphyton taxonomic composition. A state-wide project supported by Virginia Department of the Environment (VADEQ) analyzed the scrapings for a number of parameters at locations at or near benthic macroinvertebrate and water quality stations. Most of the periphyton analysis done for this study was focused on their results.

An array of data types can be matched with the response variables' sampling events and used to discern confounding factor effects and investigate potential nutrient thresholds. Certain water quality measurements are made as a matter of course when biological monitoring programs collect biological samples. These include pH, conductivity, dissolved oxygen, turbidity, and water temperature. Often, the same programs simultaneously collect samples for laboratory analysis of total nitrogen (TN), total phosphorus (TP), their various dissolved and particulate forms, total suspended solids (TSS), and dissolved organic carbon (DOC). Similarly, habitat condition evaluations are made when most macroinvertebrate samples are collected. These could include bank stability, bank vegetation, riparian vegetation, embeddedness, sedimentation, cover, channel alteration, bank and wetted widths, and riffle frequency. Additional information extracted from GIS layers can be matched with the biological sampling locations. The most useful geo-spatial information with respect to this study was ecoregion type and Strahler stream order.

Obtaining and assembling the state agency data sets into common formats was a large part of this study. The data sources and applications are discussed in more detail in each of the following chapters.

3. Conceptual Framework

A major objective of our analyses was to minimize to the extent possible the influence of environmental factors that confound biological responses to nutrients. Several key concepts create a framework of well supported hypotheses that explain and integrate the body of knowledge about stream and river ecosystems. This conceptual framework guided the analyses. The "River Continuum" and "Physiographic Region" concepts describe important natural causes of variation in aquatic communities. They were used to divide the analysis datasets into more homogeneous groups to be analyzed separately. The third concept, which we are calling the "Abiotic Condition Gradient," deals with physical and chemical aspects of stream and river environments that at stressful levels overwhelm biological responses to nutrients. Sampling events associated with stressful levels of some physical and chemical parameters were either "filtered" (removed) from the analysis datasets or incorporated into them as distinct classes. This approach of categorizing and filtering data in order to minimize confounding effects was possible because information about important confounders is available for many sampling events, and sample sizes and ranges of environmental conditions in the resulting data groups are sufficiently large.

Exploratory analyses confirmed the confounding effects of several environmental factors on phytoplankton, periphyton, and macroinvertebrates nutrient responses. The RPART method was used to investigate how to minimize or remove those confounding effects. RPART models consisted of some measure of the aquatic community as the dependent variable and TP, TN, and the metrics of several potentially confounding environmental factors as independent variables. Model runs produced binary trees with information about which metrics most effectively subdivide the data into increasingly homogeneous groups and which metric values act as the best breakpoints for splitting the data. If tree splits are best achieved with the environmental factors instead of TP or TN, the biological metric's relationship to nutrients is considered to be confounded, or less important than its relationships to the other environmental factors. Splitting and filtering the data sets and rerunning the RPART models eventually produces binary trees with splits best achieved with TP and/or TN (and surrogates for light attenuation in the case of phytoplankton), which indicates the influences of the other environmental factors have been minimized or removed.

The first two sections below discuss how natural, large-scale factors described in the River Continuum and Physiographic Region concepts are thought to affect aquatic communities. The specifics of how the two concepts were used to subset the phytoplankton, periphyton and macroinvertebrate data are described in the relevant report chapters. The third section below addresses how important physical and chemical parameters of the immediate habitat are expected to affect biological communities. These parameters often experience the greatest anthropogenic impacts. The relative importance of each parameter to the phytoplankton, periphyton and macroinvertebrates and the parameter values filtered from the analysis datasets are described in the relevant report chapters.

3.1 Physiographic Region Concept

The ecoregion framework developed by Omernik and colleagues was the basis for dividing the study region into units with similar soils, natural vegetation, climate, structural/bedrock geology, land use, hydrology, and glacial history. Ecoregions are specifically designed to provide a classification framework for ecological analysis of spatial datasets (Omernik 1995, Woods *et al.* 1999). The classifications are hierarchical in nature, with the uppermost Level I dividing the continental United States into 15 major regions and lowest Level IV characterizing hundreds of units in fine resolution. Most of the monitoring

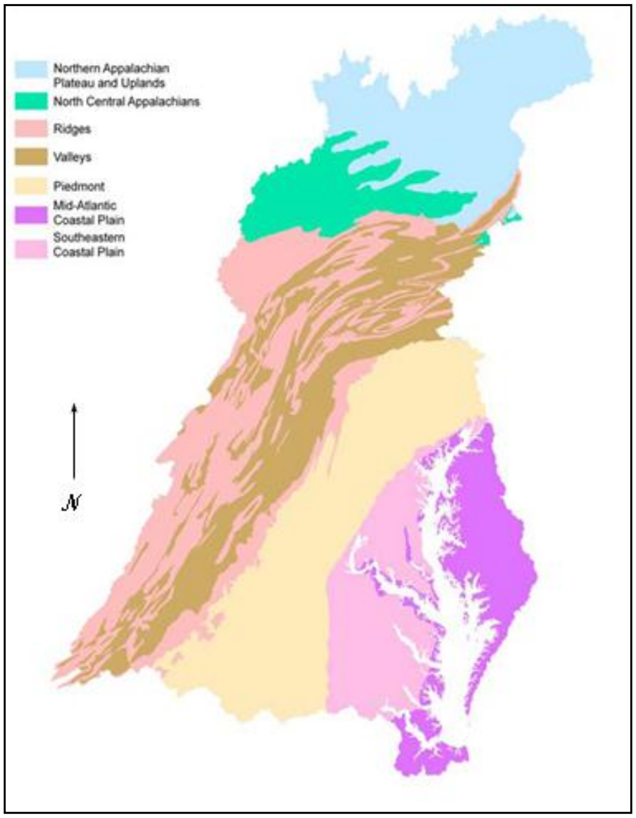


Figure 2. Bioregion framework in the Chesapeake Bay basin (Figure 1 in Buchanan *et al.* 2011).

data analyzed in this study were collected in the Chesapeake Bay basin which falls in the Eastern Temperate Forest and Northern Forest Level I ecoregions. Within the basin, the Level I ecoregions are divided into 5 Level II ecoregions, then 13 Level III ecoregions, and finally 39 Level IV ecoregions.

Dividing the study data into the 39 Level IV ecoregion groups would create subsets of the data with sample sizes too small to analyze with confidence. Foreman *et al.* (2008) and Buchanan *et al.* (2011) tested Level IV ecoregions as a framework for classifying stream macroinvertebrate communities and found that Level IV ecoregions could be aggregated into units somewhat similar to Level III ecoregions but retaining division where true differences in the macroinvertebrate fauna occurred. The resulting "bioregions" that overlap the state of Maryland were used as a testable classification system in this study. They are Piedmont, Ridges, Valleys, Mid-Atlantic Coastal Plain, and Southeastern Plain (**Figure 2**). If the exploratory RPART models showed strong, primary splits by bioregion, the data were subsequently analyzed by bioregion. If model runs indicated the biological response was not sensitive to bioregion, the data were grouped to increase sample sizes.

3.2 River Continuum Concept

The longitudinal gradient of environmental changes in free-flowing waters as they move from small headwater streams to large rivers is one of the most important causes of change in aquatic communities and can confound evidence of anthropogenic disturbance. The environmental gradient and associated biological changes were first synthesized as the River Continuum Concept by Vannote *et al.* (1980); details of the concept are being refined but the fundamental aspects are in place (see Wetzel 2001). In undisturbed systems, structural and functional attributes of communities adapt to the physical gradient in predictable ways by conforming to the energy inputs, hydrodynamics, and physical properties of the river system where they are located. This creates a longitudinal shift in the dominant producers from periphyton and vascular plants to phytoplankton in the downstream direction, and a similar shift in the dominant macroinvertebrate consumers from shredders to grazers and collectors (**Figure 3**).

Stream size as expressed by Strahler or watershed size effectively groups key attributes of undisturbed lotic environments and communities. Headwaters and streams (Strahler order 1 – 3) are shaded to a large degree by overhanging riparian vegetation which suppresses photosynthesis but contributes large amounts of coarse particulate organic matter in the form of leaf litter. The organic matter, decomposed by bacterial and fungal biofilms, supports the predominantly heterotrophic food webs of headwater streams. Shredders and collectors dominate the headwater macroinvertebrate community. As stream size increases, the amount of shading decreases (**Figure 4**) and the corresponding rise in incident light favors growth of periphyton and rooted aquatic plants. Primary production is maximized in large streams and mid-sized rivers (Strahler order 4 – 6) although upstream processing contributes some fine particulate organic matter. Collectors and grazers dominate the macroinvertebrate community. In large, deep rivers, phytoplankton replace periphyton and rooted aquatic plants as the major primary producer. The river system remains autotrophic overall if water transparency is good, but gradually falls back to heterotrophy if transparency is poor. Filter feeding collectors usually dominate macroinvertebrate communities in large systems.

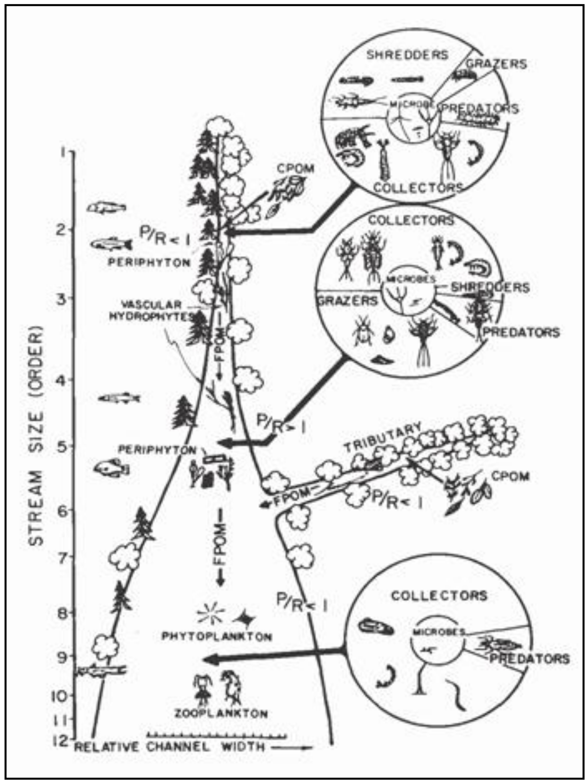


Figure 3. Diagram of the relationship between stream size and the progressive shifts in structural and functional attributes of lotic communities (from Vannote *et al.*, 1980). CPOM, coarse particulate organic matter; FPOM, fine particulate organic matter; P/R, ratio of gross primary productivity to community respiration.

A Strahler stream order assignment was obtained for each sampling location from the National Hydrography Dataset (1:100,000) stream layer. Strahler order in the Potomac River's Piedmont, Ridges, and Valleys bioregions is significantly related to watershed size. A comparison of 700+ delineated watersheds from these bioregions also demonstrated good overlap between Strahler order and the Northeast Aquatic Habitat Classification System (NEAHCS) for streams and rivers developed by Olivero and Anderson (2008) for the thirteen northeastern states (**Table 1**). In this study, we adopt the general terminology of "stream" for Strahler orders 1-4 and "rivers" for orders 5+.

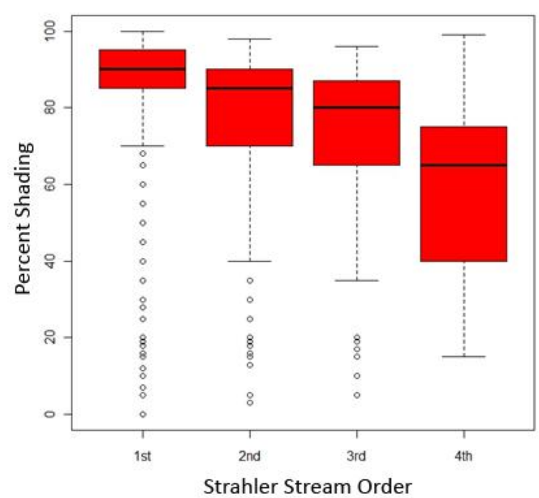


Figure 4. Percent of water surface shaded by riparian vegetation at relatively undisturbed sites (riparian buffer scores 16 - 20) in the Chesapeake Bay basin's Piedmont, Ridges, and Valleys bioregions.

Strahler order was included as an independent variable in exploratory RPART model runs for macroinvertebrate and periphyton. The results were used to classify data into distinct groups for the nutrient response analysis. Phytoplankton data associated with Strahler orders 1-4 were initially excluded from the nutrient response analyses. They were later added back to the coastal plain data sets when it became apparent Strahler order was not confounding the analysis in those bioregions and the larger counts aided the analysis.

Table 1. Comparison of Strahler order and Northeast Aquatic Habitat Classification System (NEAHCS) size classes in the Piedmont, Ridge, and Valley bioregions of the Potomac River basin. The Potomac River above Little Falls is a Strahler order 7 river; its watershed is 10,700 sq. mi. which makes it a NEAHCS size classes 5 "Great River."

Strahler order…	… and NEAHCS size class	…match in this % of comparisons	NEAHCS upstream drainage area (sq. mi.)	NEAHCS description	Description used in this report
1	1a	91.3%	0 - < 3.861	Headwaters	Stream
2 - 3	1b	63.5%	3.861 - <38.61	Creeks	Stream
4	2	73.6%	38.61 - <200	Small Rivers	Stream/Small River
5 - 6	3	78.9%	200 - <3,861	Medium Rivers	River
7+	4+		3,861+	Large/Great Rivers	River

3.3 Abiotic Condition Gradient Concept

As mentioned in the report introduction, simple regressions rarely reveal strong relationships between a biological response variable and a nutrient gradient. Instead, a classic wedge-shaped pattern emerges (**Figure 1**) where other environmental factors are thought to control the biological community. An important step in distinguishing the impacts of a nutrient on stream and river communities is to separate to the extent possible responses to nutrients from responses to stressful levels of other in-stream factors.

The responses of biological communities and taxa to ecologically important chemical and physical factors have been established through observation and experimentation (see summaries in Wetzel 2001, Thorp and Covich 2001, Lampert and Sommer 1997, Merritt and Cummins 1996, Allen 1995, Gordon *et al.* 1992, and Ward 1992). A certain level of disturbance and stress is normal, even in high quality environments. Organisms usually survive brief periods of extreme water quality conditions; they require a narrower range of conditions to function adequately over time, and still narrower ranges to successfully reproduce. Dissolved oxygen, pH, and conductivity are three important water chemistry parameters for which organisms have variable tolerance limits depending on whether survival, maintenance, or reproduction metrics are being measured. Suspended sediment levels are normally elevated in flowing waters because of turbulence, and the phytoplankton, submerged aquatic plants, and periphyton found in streams and rivers have adapted to these conditions; however, very high concentrations for prolonged periods impairs photosynthesis and causes physiological stress in plants. Tolerance limits for structural disruptions in habitat are similarly variable. Organisms can survive some scouring and habitat alteration but health and successful reproduction requires a certain degree of habitat stability and good quality. Behavioral responses to other factors can result in organisms living closer to their tolerance limits rather than at their optimum levels. For example, an organism may frequent sub-optimal habitats in order to avoid a predator or reduce competition.

The flexibility and diversity of tolerance limits to chemical and physical stress makes it difficult to decide which records to filter from an analysis and still maintain a data set that holistically represents undisturbed environments exposed to a broad nutrient gradient—in other words, environments where nutrient responses are not significantly confounded by other stressors. Furthermore, there are environmental factors that should not be filtered from the data because of their fundamental importance in cellular processes. For example, light energy in deep water should be considered together with plant nutrient responses because of its role in photosynthesis and nutrient assimilation. Finally, cellular processing of nitrogen and phosphorus—as well as other nutrients—and production of new tissue is affected by the forms and proportions of nutrients available for uptake and consumption. Resolving nutrient responses to this level of detail using monitoring program results is not practical.

The "binning approach" used in this study filters and then groups data records into distinct nutrient categories that also reflect natural variability introduced by other abiotic factors in relatively undisturbed environments. It effectively creates a nutrient gradient in the context of other abiotic conditions where those conditions do not overwhelm (confound) nutrient responses of the biological community. Several methods were used to decide if and where to filter data:

> 1) Macroinvertebrate records associated with dissolved oxygen and pH levels that appear to fail water quality standards were filtered from the data sets. Some allowance was made for exceedances of the standards, so the parameter thresholds used to filter the data varied depending on bioregion and do not necessarily match Maryland water quality criteria.[1]
>
> 2) If exploratory RPART model runs showed consistent splits on a chemical or physical parameter, with approximately the same breakpoint value identified in the splits, data records on the stressful side of the parameter breakpoint are removed. Examples are high levels of conductivity and low stream habitat scores which are consistently related in RPART models with poor macroinvertebrate communities.
>
> 3) If a parameter is known to have a dominant functional role, it is incorporated into the different nutrient classes or bins created for the nutrient response analyses. An example is light, and specifically light attenuation, which impairs photosynthesis in phytoplankton and periphyton.

"Bins" or categories were then derived from the recurring breakpoints identified in RPART analyses on the filtered data.

[1] Maryland COMAR 26.08.02.03-3 states that dissolved oxygen concentrations may not be less than 5 mg/liter at any time in non-tidal warm waters (Uses I, I-P, IV, IV-P) and may not be less than 5 mg/liter at any time, with a minimum daily average of not less than 6 mg/liter, in non-tidal cold waters (Uses III, III-P). Maryland declares a pH impairment if pH values in 10% or more of samples from an 8-digit stream watershed are >8.5 or <6.5, and violations cannot be traced to naturally occurring conditions (MDE 2008).

4. Phytoplankton

Eutrophication in large rivers can be expressed as "blooms" of free-floating algal cells, or phytoplankton. River phytoplankton populations are seeded by cells dislodged from the nearshore periphyton or flushed from upstream (lake or reservoir) phytoplankton communities. Riverine algal blooms occur when the river is sufficiently wide and deep to create pelagic conditions, flow velocities are slow enough to allow algal cells to accumulate, and nutrient concentrations are high. A negative consequence of algal blooms can be hypoxia, or the depletion of oxygen, related to strong night-time respiration by living cells and/or the bacterial decomposition of dead cells. Algal blooms also create costly taste and odor problems for water suppliers and some blooms species such as *Microcystis aeruginosa* can produce toxins that sicken or kill humans and livestock who contact or drink contaminated water.

The phytoplankton analysis was designed to first identify the influence of important confounding factors and then account for them by filtering and grouping (classifying) the data so as to track the effects of the factors across a range of nutrient conditions. Nutrient breakpoints for each subset of the data were then determined above which degradation or impairment occurs at an increasing frequency. The most important confounding factor for phytoplankton was expected to be light—or more specifically light attenuation—in the water column of the river (Wetzel 2001). When suspended particles and dissolved substances attenuate incident light energy, photosynthetic gains are lost to increasing respiration costs and phytoplankton become physiologically stressed. One mechanism phytoplankton use to counter low light stress is to raise the chlorophyll *a* content of their cells (e.g., Kirk 1994) which in turn enhances their capacity to bloom if and when water movements carry them into more favorable light conditions. Flow velocity is another important confounder. For example, slow moving streams in the flatter coastal regions can be more conducive to phytoplankton population growth than the faster moving, higher gradient streams in regions above the Piedmont fall-line where downstream displacement significantly undercuts population growth.

For this data analysis and report, a chlorophyll *a* (Chla) concentration of 30 µg/liter was used to define blooms. The concentration is currently used as an endpoint for assessing Maryland lakes and reservoirs. Instantaneous exceedances of 30 µg/liter trigger increased scrutiny during routine data analysis of monitoring results and managerial inquiry during TMDL analysis and development. The study's analysis approach lends itself to the application of other Chla thresholds and criteria, and these can be explored in the future.

4.1 Analysis methods

Relationships between phytoplankton biomass and nutrient enrichment in free-flowing rivers were explored with water quality data collected by MDDNR (Core Trend Program), MDE (multiple programs), and DE DNREC (Biological Monitoring Program). Chlorophyll *a* (µg/liter) served as the surrogate for phytoplankton biomass. Nutrient enrichment was quantified from TP and TN concentrations (mg/liter). Turbidity (NTU) was used as the variable used to characterize the light environment. Technically, turbidity measures light scattering caused by suspended particles and not attenuation which is the loss of light energy due to absorption. Dissolved organic carbon (DOC, mg/liter) was an additional required variable in the Coastal Plain region where it served as a measure of the spectral shifts and additional light attenuation caused by the naturally high concentrations of dissolved organic compounds ("blackwater"). DOC was also tested as a confounding factor in the non-Coastal Plain regions although levels there are about half of those in the Coastal Plain. The MDDNR stations were with few exceptions sampled year round on a monthly basis in each year between 1986 and 2006. The MDE stations were often sampled year round but only for 3 to 8 of the 12 years between 1998 and 2009. Water quality at the DE DNREC stations were sampled intermittently, usually in spring, summer or autumn, between 2003 and 2008.

4.1.1 Classification

Data were divided *a priori* into Coastal Plain and non-Coastal Plain groups. At first, only monitoring stations on rivers with a Strahler order assignment of 5 or larger were considered in each physiographic group (**Figure 5**). The River Continuum Concept (Vannote *et al.* 1980) makes the case that phytoplankton populations are most suited for these large rivers. Photosynthesis is not suppressed by riparian vegetation shading, mid-channel waters are in theory too deep for significant competition from periphyton and rooted aquatic plants, and downstream displacement is slow enough to allow phytoplankton populations to grow. Of the monitored non-tidal rivers in the non-Coastal Plain region, the Potomac River mainstem attains Strahler order 7 and seven Maryland tributaries attain orders 5 or 6. Of the monitored non-tidal rivers in the Coastal Plain, none are larger than Strahler order 5 and only three attain Strahler order 5 for significant distances: Choptank, Nanticoke and Patuxent (**Table 2**).

The non-Coastal Plain group has monitoring data for 58 stations classified as Strahler order 5-7 (**Table 2**). All the stations are located in the Piedmont, Ridge, and Valley bioregions in Maryland. Rivers in this group were not further separated by bioregion because they all cross more than one bioregion before entering their Strahler order 5-7 reaches, making bioregion a useless classification. A total of 4,103 samples had the requisite chlorophyll *a*, TN, TP, and turbidity data. A subset of 745 samples also had DOC data.

In the Coastal Plain group, only 206 sampling events from 13 stations were found in the rivers attaining Strahler order 5. Strahler order classification in this region was eventually dropped when it became apparent that nutrient breakpoints derived from the region's Strahler order 5 rivers also successfully binned water quality conditions and phytoplankton responses in its streams and small rivers. The Coastal Plain's flatter land surface and slower stream velocities with their associated longer residence times allow water column algal blooms to form in streams and rivers of all sizes. Furthermore, extensive ditching to channel and drain surface flows has significantly changed the historic stream flow patterns, making Strahler order a somewhat unreliable measure of branching complexity and stream size.

The expanded Coastal Plain group consisted of data from 348 stations in Delaware and Maryland streams and rivers. There were 1,911 sampling events in the Mid-Atlantic Coastal Plain ("eastern shore") of which 93% were from Strahler order 1-4 reaches. There were 1,405 sampling events in the Southeastern Plain ("western shore") of which 94% were from Strahler order 1-4 reaches. The Coastal Plain data were analyzed separately by bioregion

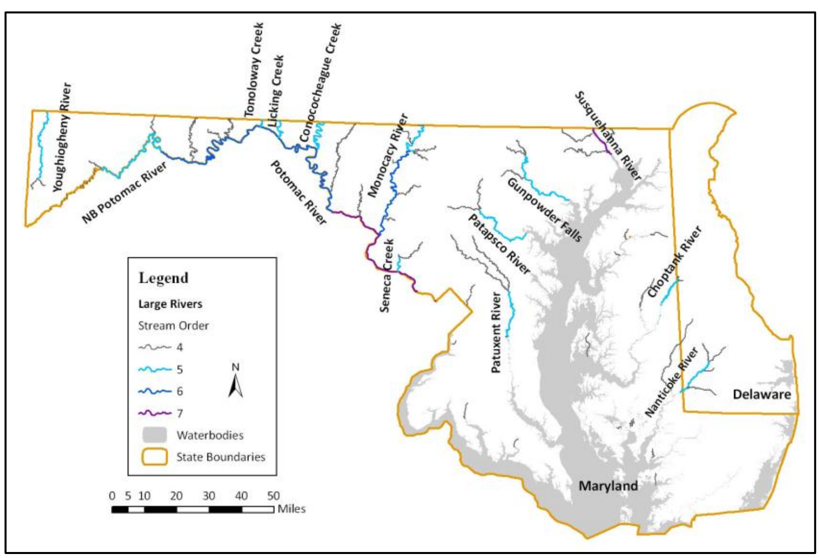


Figure 5. Strahler order 5-7 rivers used in the data analysis.

Table 2. Maryland medium and large rivers. The down- and upstream boundaries of each river reach of Strahler order 5 or greater are indicated in river miles (RM) as measured [1]from the mouth of the river's confluence with a larger river, [2]from tidal waters, or [3]from the Maryland state boundary. [4]Sampling locations in Delaware's Nanticoke River Strahler order 5 reach are included. Stations were counted if they had sampling events with the requisite TN, TP, turbidity and chlorophyll *a* data (and DOC in the Coastal Plain).

Non-Coastal Plain	# Stations	Strahler Order	Reach (RM)
Conococheague Creek[1] MD	7	5	0 – 21.7
Gunpowder River[2] MD	9	5	3.6 – 28.4
Monocacy River[1] MD	13	5-6	0 – 57.5
Patapsco River[2] MD	6	5	14.8 – 34.7
Potomac North Branch River[1] MD	7	5-6	0 – 51.4
Potomac River[2] MD	7	7	118.4 – 238.6
Tonoloway Creek[1] MD	4	5	0 – 1.5
Youghiogheny River[3] MD	5	5	91.8 – 115.9
Mid-Atlantic Coastal Plain			
Choptank River[2] MD	2	5	62.6 – 65.4
Nanticoke River[4] MD, DE	7	5	31.2 – 45.5
Southeastern Coastal Plain			
Patuxent River[2] MD	4	5	54.6 – 61.3

because the Mid-Atlantic Coastal Plain bioregion has significantly higher levels of DOC ($p<0.001$), and hence a greater degree of light attenuation and spectral shifting, than the Southeastern Plain bioregion.

4.1.2 Chlorophyll a adjustment of TN, TP, and turbidity values

Phytoplankton are a component of the turbidity readings and TN and TP concentrations measured in streams and rivers. Therefore, water column chlorophyll *a* relationships to these values are confounded to some extent by self-correlation. A simple empirical method was used to conservatively estimate the proportions of TN, TP, and turbidity attributable to phytoplankton and subtract those amounts from the measured values of each parameter. The adjusted values represent to a large degree the non-algal components of each parameter. Thus, an analysis comparing Chla to adjusted TN, TP and turbidity values is less circular than one comparing Chla to unadjusted values.

For over 30 years, researchers have pointed out the difficulty of separating light attenuation caused by algal and non-algal constituents, as well as the potential to misrepresent trophic state and recommend inappropriate management strategies if the non-algal constituents of light attenuation are not considered (e.g., Lorenzen 1980, Megard *et al.* 1980). Although turbidity measures light scattering and absorption rather than just light absorption, the difficulties are the same. To differentiate the phytoplankton component of turbidity from other particles, the individual scattering/absorption properties of phytoplankton and all non-algal constituents in the water column would need to be known. Similar difficulties exist in distinguishing the phosphorus and nitrogen contents of living phytoplankton cells from non-algal water column constituents. These include the various dissolved forms of phosphorus and nitrogen as well as the phosphorus and nitrogen contained in bacteria, detritus (includes dead phytoplankton cells), and zooplankton.

A simple graphical approach was used to remove at least some of the algal component from turbidity, TN, and TP values and minimize the confounding effect of self-correlation. Turbidity, TN, and TP values were plotted against their corresponding chlorophyll *a* measurements and lines bounding the lowest TN, TP and turbidity values at each chlorophyll concentration, excluding probable outliers, were established by eye using a linear regression through the lowest points (**Figure 6**). Distances below these boundary lines to the x-axis conservatively represent the amount of algal matter comprising each parameter at the corresponding chlorophyll *a* concentration.

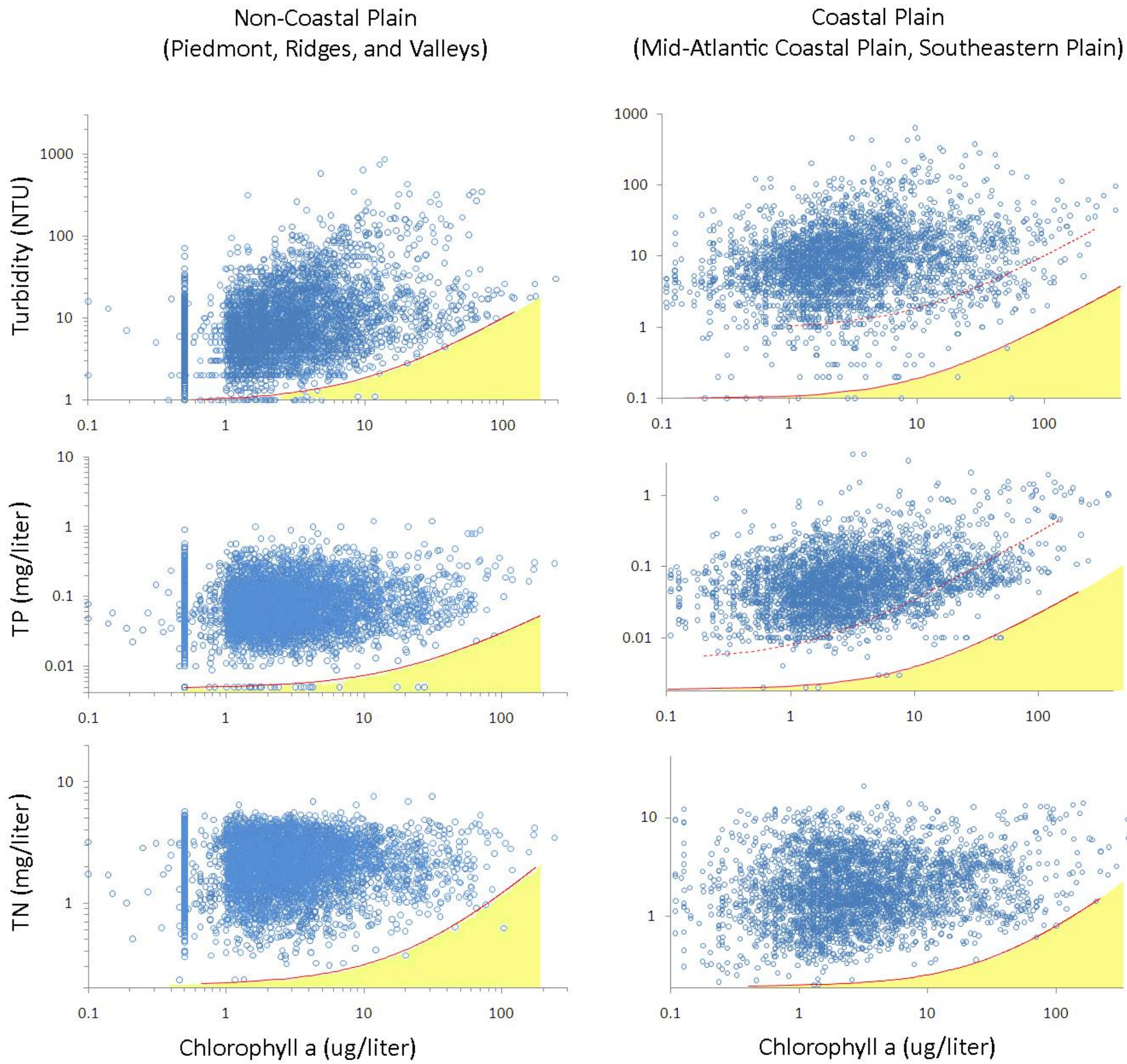


Figure 6. Chlorophyll adjustment of turbidity, TP, and TN values. Red solid lines indicate the equations used to calculate the algal component for each parameter at a specific chlorophyll *a* concentration (see text for details). Distance from the red line to the x-axis (shaded area) is conservative estimate of the algal component of turbidity, TP, or TN at the corresponding chlorophyll *a* concentration. Distance above the line to an individual data point represents the non-algal components. The non-Coastal Plain lines are superimposed on the Coastal Plain data as a dashed line to illustrate the differences found between the Coastal and non-Coastal Plain equations for turbidity and TP. The TN lines for the two groups overlap closely.

Using the boundary line equations, the algal-related amounts were subtracted from each record's turbidity reading and TN and TP concentrations to obtain rough estimates of the non-algal amounts:

Non-Coastal Plain (Piedmont, Ridges, and Valleys)

- Adjusted Turbidity = AdjTurb = [Turbidity] - ((0.091*[Chla]) + 0.955)
- Adjusted TP = AdjTP = [TP] - ((0.0003*[Chla]) + 0.0049)
- Adjusted TN = AdjTN = [TN] - ((0.01*[Chla]) + 0.2114)

Coastal Plain (Mid-Atlantic Coastal Plain, Southeastern Plain)

- Adjusted Turbidity = AdjTurb = [Turbidity] - ((0.0091*[Chla]) + 0.098)
- Adjusted TP = AdjTP = [TP] - ((0.0002*[Chla]) + 0.0019)
- Adjusted TN = AdjTN = [TN] - ((0.006*[Chla]) + 0.1922)

Turbidity units are NTU; TP and TN units are mg/liter. The data reveal different adjustment equations for turbidity and TP in the Coastal and non-Coastal Plain groups (note dashed lines in **Figure 6**) but similar adjustment equations for TN.

The turbidity, TN, and TP adjustments are not precise. Ideally, equations would be developed for groups or classes of free-flowing waters with similar properties. The non-Coastal and Coastal Plain groupings do this to some extent. For example, significantly higher concentrations of DOC characterize the latter group. However, the rivers and streams in each group have fairly broad ranges of other potentially confounding variables such as conductivity and alkalinity, and have varying proportions of particulate and dissolved nutrient forms. In addition, the density of the data points along the gradient of chlorophyll *a* concentrations will affect the position and shape of the adjustment equation line and ultimately the level of confidence in the adjustment. As can be seen in **Figure 6**, data points are relatively scarce at very high chlorophyll *a* concentrations. Finally, some of the very low chlorophyll *a*, turbidity and TP concentrations in the non-Coastal Plain group have been rounded to detection limits: 0.5 µg/liter for Chla, 0.05 mg/liter for TP, and 1 NTU for turbidity. Despite the adjustment method's inherent weaknesses, it does reduce to a large extent and in a consistent manner some of the confounding effects of chlorophyll self-correlation. Adjusted turbidity, TN, and TP values were used in the phytoplankton analyses to identify nutrient breakpoints. The breakpoints were then transformed back to unadjusted values (see 4.3.3 below) so both adjusted and unadjusted values can be considered.

4.1.3 Recursive partitioning

The RPART routines that construct regression trees were used to explore phytoplankton relationships to nutrients and light in the non-Coastal Plain, Mid-Atlantic Coastal Plain (MACP), and Southeastern Plain (SEP) physiographic groups. In each group, an RPART model was constructed and tested with up to four different minimum split size "rules" (n = 20, 30, 40, 50). The RPART model tested on 4,103 records in the non-Coastal Plain region, or the combined Piedmont, Ridges, and Valleys (PRV) bioregions, was:

Chla ~ AdjTurb + AdjTP + AdjTN

The RPART model tested on 1,783 records in the Mid-Atlantic Coastal Plain and 1,472 records in the Southeastern Plain was:

Chla ~ AdjTurb + AdjTP + AdjTN + DOC

This latter model was also tested on 745 records from the non-Coastal Plain that had DOC data.

Each model run of the RPART program produces a decision-tree identifying the "primary" splits at each node of the tree, or more specifically the value of each independent (competing) parameter that best splits the data at that node. All of the independent parameters are listed according to how well they reduce the variance of the dependent variable in the two resulting branches (best, second best, etc.). The parameter that yields the best split is shown in the resulting tree diagram. "Surrogate" splits are also produced which identify other splits that classify the same data points in the same way. All of the primary and surrogate splits are considered potential breakpoints for each independent variable. The values of all the primary and surrogate splits for each independent variable were extracted from the RPART "summary" output, combined, and sorted. Values that were identified as the best primary split at each node and/or identified multiple times as a primary or surrogate split were flagged. Flagged values became the candidate breakpoints used to construct the classes or "bins" of water quality parameters.

4.1.4 Binning

The objective of the binning step was to use the candidate breakpoints in AdjTurb, AdjTP, and AdjTN (and DOC in the Coastal Plain) to create water quality bins with unique environmental conditions that have distinct biological responses grading towards degradation. Selected candidate breakpoints serve as boundaries of the bins. Each bin is an ensemble of a set of nutrient conditions and the associated levels of one or both confounding light factors. Bins are established through an iterative process aimed at finding those combinations of parameter ranges that result in distinct phytoplankton responses (chlorophyll *a* concentrations). Several "high" and "low" bins are created first using combinations of the higher and lower candidate breakpoints, respectively, until adequate sample sizes and consistent responses occur at both ends. Various combinations of the independent variables are then used to create contrasting bins between the high and low ends. The process can be greatly influenced by the size and character of the data set. For example, a "high" bin may not be possible to create in a physiographic region with relatively few records having high turbidity and high TP and high TN. The large water quality data sets used in the phytoplankton RPART analyses and binning process typically were able to populate bins at the both ends of the spectrum of conditions.

Seasonal differences were investigated after the RPART and binning analysis had identified nutrient response breakpoints. Seasonal changes in temperature and light intensity can be expected to affect phytoplankton growth rates and the frequency of algal blooms but not the breakpoints at which nutrient and light conditions favor bloom formation.

4.2 Exploratory analysis results

Exploratory analyses were done on the selected Maryland and Delaware non-tidal monitoring records, each having measurements for chlorophyll *a*, TN, TP, and turbidity (and DOC data in the Coastal Plain group). The analyses were intended to confirm the choice of analysis approach and guide the selection of nutrient breakpoints or thresholds. Stations located on river reaches classified as Strahler order 5 or greater are considered rivers. Those located on reaches classified as Strahler order 1-4 are considered streams although technically Strahler order 4 are small rivers in the NEAHCS system (**Table 1**).

Statistical summaries of Chla concentrations in the selected data set show that Maryland's river stations may not experience algal blooms often (**Table 3** below and **Appendix A Table 1**). Just 85 of the 4,103 events in the non-Coastal Plain (2.1%) and 5 of the 128 events (2.4%) in the MACP (3.9%) exceeded Chla concentrations of 30 µg/liter ("blooms"). None of the 78 sampling events in the SEP rivers experienced blooms. In this selected data set, the two MACP rivers—Choptank and Nanticoke—appear to have distinctly higher amounts of TN to TP. A look at the nutrient concentrations indicates TP is comparable across the three regions, so this higher ratio in the MACP rivers is due to an excess of nitrogen relative to phosphorous rather than to very low phosphorus concentrations. Turbidity, one of two parameters

Table 3. Water quality characteristics of medium and large rivers (Strahler order greater than or equal to 5) in the non-Coastal Plain (Piedmont, Ridges, Valleys bioregions combined), Mid-Atlantic Coastal Plain (MACP), and Southeastern Plain (SEP). Percentage in parenthesis in the Chla column is the frequency that Chla equals or exceeds 30 µg/liter in the data record of each region.

Region	Statistic	DO mg/liter	Conduc-tivity umhos/cm	pH PSU	Turbidity NTU	DOC mg/liter	TSS mg/liter	TN mg/liter	TP mg/liter	TN:TP ratio	Chla µg/liter
Non-Coastal Plain	5%ile	6.5	168	7.2	2.0	1.25	0.9	0.85	0.016	10.0	0.50
Chla n = 4,103	25%ile	8.0	240	7.6	4.0	1.93	3.1	1.66	0.038	21.2	1.19
n with DOC = 745	median	9.7	294	7.9	7.1	2.47	7.0	2.39	0.065	34.8	2.09
	75%ile	11.9	367	8.2	14.0	3.23	15.0	3.33	0.112	57.3	4.19
	95%ile	13.8	477	8.6	45.0	5.58	49.0	4.95	0.250	137.0	16.31
											(2.1%)
Mid-Atlantic Coastal	5%ile	5.6	108	5.9	2.0	2.20	2.4	1.51	0.010	16.6	0.25
Plain	25%ile	6.4	131	6.6	5.0	3.00	3.1	1.95	0.034	33.6	1.16
Chla n = 128	median	7.6	141	6.9	7.0	3.93	7.0	2.99	0.050	68.8	3.33
(Nanticoke and	75%ile	8.9	165	7.1	10.0	5.12	13.0	4.36	0.070	123.4	10.54
Choptank rivers)	95%ile	11.5	339	7.6	15.7	9.07	27.0	5.71	0.128	276.7	25.95
											(3.9%)
Southeastern Plain	5%ile	6.0	209	7.0	4.9	2.87	4.3	1.57	0.048	11.8	0.24
Chla n = 78	25%ile	7.0	274	7.2	8.7	3.77	8.1	1.78	0.066	20.6	0.67
(Patuxent River)	median	8.3	322	7.4	13.6	4.13	12.2	1.96	0.083	39.0	1.49
	75%ile	11.2	371	7.5	25.4	4.75	24.0	2.20	0.102	67.3	2.99
	95%ile	12.6	635	7.7	88.6	5.67	60.1	2.64	0.200	124.4	5.98
											(0.0%)

used to portray light attenuation, was distinctly higher in the Patuxent, the SEP river. The other parameter, DOC, was distinctly higher in both the MACP and SEP rivers compared to the non-Coastal Plain rivers. Conductivity and pH were significantly lower in the MACP rivers, with pH dropping below 6 PSU more than 5% of the time. Exploratory RPART model runs consistently separated the MACP and SEP rivers (and later, the MACP and SEP rivers and streams) so the two bioregions were analyzed separately for nutrient and light breakpoints.

In the non-Coastal Plain group, blooms were not recorded at ¾ of the river stations, and chlorophyll *a* concentrations were less than 7 µg/liter in ¾ or more of each station's samples. Of the 15 non-Coastal Plain river stations with more than 200 samples—and thus the highest level of confidence in estimates of algal bloom frequency—the overall frequency of algal blooms was highest at one Conococheague, four Monocacy, and five Potomac mainstem stations. Frequencies were 2.4% at Conococheague station CON0005, 2.0% - 4.4% at the Monocacy stations between Rt 140 (MON0528) and its confluence with the Potomac, and 2.7% - 3.5% at the Potomac River stations between Point of Rocks (POT1596) and Little Falls (POT118.4). These samples reach back as far as the year 1986. In a subset of data between 2000 and 2009, bloom frequencies at the same stations were 1.0% at the Conococheague station, 0.0% - 2.5% at the Monocacy stations, and 3.6% - 5.1% at the Potomac stations—or slightly lower in the Conococheague and Monocacy and slightly higher in the Potomac mainstem. About 2/3 of the blooms occurred in summer, but a quarter occurred in spring and a few in autumn and winter. Overall, 2.07% of the 4,103 samples in the analysis data set were experiencing an algal bloom at the time of sample collection.

Coastal Plain river samples with the requisite water quality parameters were only found for 2003 - 2007. Bloom frequencies at the 13 individual river stations cannot be calculated with confidence because of the small sample sizes at most stations. Patuxent stations PXT0613 and PXT0561 and Choptank station CHO0626 have the largest number of samples with 35, 23, and 31 samples, respectively. Chlorophyll *a* concentrations greater than 30 µg/liter were not found in the Patuxent, but occurred in 3.2% of the samples from the Choptank. There are just 239 samples from Maryland rivers in the two coastal plain bioregions.

The Coastal Plain data set increases to 3,255 samples with the addition of data for Maryland streams and Delaware streams and rivers in the Chesapeake Bay basin. Approximately 10.38% of the 1,783 stream and river sampling events in the Mid-Atlantic Coastal Plain and 1.02% of 1,472 sampling events in the Southeastern Plain had chlorophyll *a* concentrations greater than 30 µg/liter. It is difficult to identify with confidence the streams and rivers that most frequently experience algal blooms because Coastal Plain stations have not been repeatedly sampled over a long time period. Of the streams and rivers with 20 or more samples, the following had a greater than 10% bloom frequency: Wrights Branch, Middle Neck Branch, Transquaking River, Chicamacomico River, Fowling Creek, Leonard Mill Pond, Nanticoke River, North Prong Leonard Pond Run, Wicomico River, and an unnamed tributary to Muddy Creek. All were in the Mid-Atlantic Coastal Plain. Approximately 70.5% of the blooms occurred in summer, 20% in autumn and winter, and 9.5% in spring.

Another analysis compared multiple parameters along a longitudinal cross-section of each large river reach and confirmed that no river shows a consistent change in Chla when compared to individual changes in AdjTP, AdjTN, or AdjTurb (correlation). A high average TP concentration does not dictate frequent algal blooms; neither does a high average TN concentration. These findings point out one of the weakness of regression analyses based on the assumption that an increase in a single nutrient will result in an increase in Chla. Covariance between Chla and a nutrient factor is usually not found if another factor is stressing or limiting phytoplankton abundance, for example light limitation.

Appendix A provides water chemistry summaries for the individual river stations used in this study's data analysis (**Appendix A Table 2** and **3**). It also provides water chemistry summaries by Strahler order for Coastal Plain streams (**Appendix A Table 4**). In comparing the Coastal Plain rivers (**Table 3, Appendix A Table 3**) and streams (**Appendix A Table 4**), it appears that the smaller systems are more frequently affected by high TSS, turbidity, and DOC, all of which attenuate incident light. Nitrogen is roughly comparable; phosphorus tends to be higher in small streams.

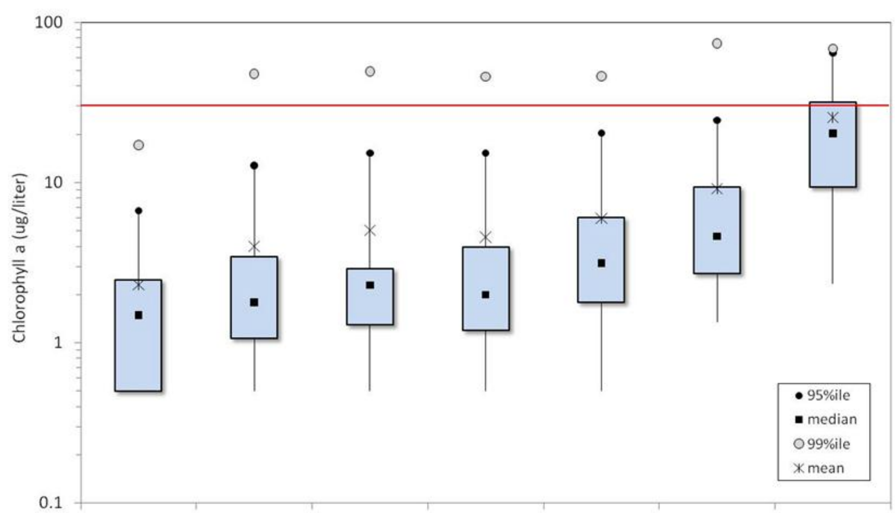
4.3 Nutrient binning results

Recursive partitioning applied with R software identified multiple breakpoints for AdjTP, AdjTN, and AdjTurb in the river data from the combine Piedmont, Ridges, and Valleys (PRV) region, and for these three parameters and DOC in the two Coastal Plain bioregions. Combinations of breakpoints were iteratively tested to create a range of distinct water quality conditions grading toward a higher frequency of algal blooms.

4.3.1 Piedmont, Ridges, and Valleys rivers

As mention above, rivers in this group were not separated by bioregion because they all cross more than one bioregion before entering their Strahler order 5-7 reaches, making bioregion a meaningless classification. Seven abiotic condition bins could be resolved in the combine Piedmont, Ridges, and Valleys bioregions (**Figure 7**). Criteria for all three independent variables—AdjTurb, AdjTN, and AdjTP—had to be met before a sampling event observation could be assigned to a bin. Bins 1 and 2 had the lowest adjusted TN and TP concentrations and the best water clarity, expressed as low adjusted turbidity. These bins represent the best water quality conditions with respect to the three water quality parameters. Bins 6 and 7 had the highest adjusted TP, TN, and turbidity levels and represent the most degraded water quality conditions. Bin 3 has relatively good water clarity and a mix of high and low nitrogen and/or phosphorus concentrations. Bin 5 has relatively poor water clarity and a mix of high and low nutrient concentrations. Bin 4 has intermediate water clarity and a mix of high and low nutrient concentrations.

Bins 1-4 have essentially the same Chla concentrations despite their nutrient and turbidity differences. Bin 5 shows increases in the median and upper quantiles, indicating a slight nutrient/light effect. Bin 6 experiences significantly more algal blooms than the preceding bins, and Bin 7 experiences blooms



BIN	1	2	3	4	5	6	7	
AdjTurb	<1.9	<7.2	<7.2	7.2 - <9.6	≥9.6	≥9.6	≥16.4	NTU
AdjTP	<0.160	<0.180	mixed	mixed	mixed	≥0.180	≥0.377	mg/liter
AdjTN	<3.20	<3.49 excl 1	mixed	mixed	mixed	≥3.49 excl 7	≥3.88	mg/liter
n	541	1374	445	392	1249	82	20	4103

Figure 7. Piedmont, Ridges, and Valleys bioregion binning results. Thresholds used to create the abiotic condition gradient bins are given below the graph. Red line indicates 30 μg/liter Chla. See text for details.

roughly a third of the time.

The overall frequency of bins 6 and 7 occurring in the Maryland PRV region is fairly low, about 2.5%. The frequency of algal blooms in just those two bins combined is about 10%. Short term sampling at a station resulting in fewer than 50 samples, for example, is not going to detect algal blooms with any degree of confidence. Bin 5 is much more common and the combined frequency of bins 5-7 at a station appears to covary with the overall frequency of algal blooms at that station (**Figure 8**). If a station is observed to experience bins 5-7 in more than a ¼ of its samples, it is more likely to experience algal blooms than stations experiencing fewer instances of bins 5-7.

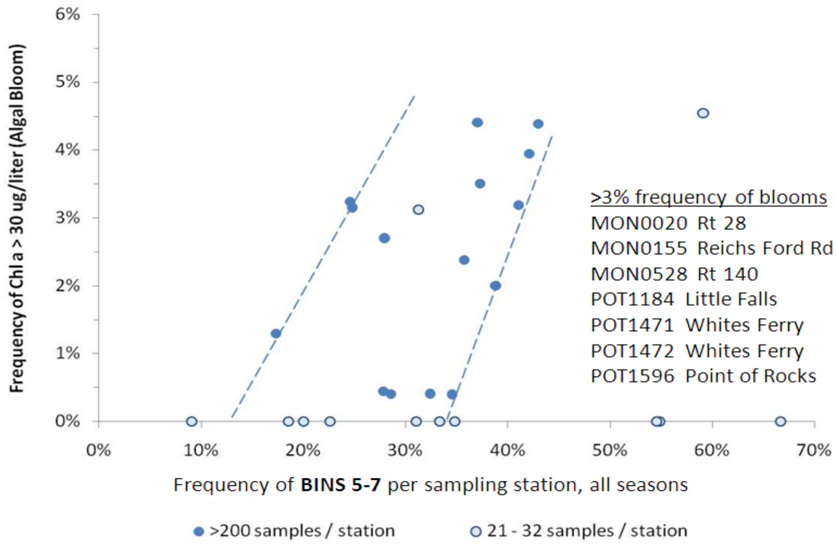


Figure 8. Relationship between frequency of bins 5-7 and frequency of algal blooms at individual stations in large rivers of the Piedmont, Ridges, and Valleys bioregions. At stations with relatively few samples (o), low estimates of algal bloom frequencies cannot be determined accurately since at least 100 samples are required to do so.

The importance of water clarity as a third factor in the bins is evident when a similar set of bins is created without AdjTurb as a factor. Bins based just on AdjTP and AdjTN do not show the smooth progression towards degradation that is seen in **Figure 8**. Algal blooms occur in all but one of the bins.

Table 4 gives for each PRV bin the median value of the AdjTurb, AdjTN, and AdjTP; the median and range of Chla concentrations; the frequency of algal blooms; the median and range of the unadjusted (observed) turbidity, TP, TN, and DOC; and the median N:P ratio calculated from the AdjTN and AdjTP.

4.3.2 Coastal Plain rivers and streams

Model runs in the RPART analyses for nutrient, turbidity, and DOC breakpoints were performed on the SEP and MACP river data. The SEP and MACP regions were analyzed together because of the low sample numbers. Bioregion was retained as an independent variable, however, and the resulting RPART trees consistently separated on bioregion in the first or second splits. A variety of RPART model runs identified a suite of breakpoints, and bins were created from the most common breakpoints.

Coastal Plain sampling events differ from non-Coastal Plain sampling events in having many more Chla measurements for stream stations. Bins identified for the rivers were applied to stream Chla data to explore the feasibility of applying the river breakpoints to streams in the Coastal Plain. The stream results were surprisingly consistent with the river results. As mentioned above, the Coastal Plain's flatter land surface and slower stream velocities appear to allow water column algal blooms to form in

Table 4. Water chemistry in bins developed for rivers in the Piedmont, Ridges, and Valleys (PRV) combined bioregions. Median values for both adjusted and unadjusted turbidity, TP and TN are given for each bin. Range: interquartile range (25th%ile – 75th%ile) of unadjusted (observed) turbidity, TP, and TN values. "Mixed" nutrient conditions have Hi P + Lo N and Lo P + Hi N. *The low sample size of this bin (n=20) increases the uncertainty in this frequency value and prevents a representative calculation of exceedence rate. **DOC was only measured in 745 of the 4103 sampling events (18.2%) so the ranges and medians given here are approximations of those in the full bins. These bins correspond to those in **Figure 7**.

BIN	1	2	3	4	5	6	7
Description							
AdjTurb	Very Low	Low	Low	Moderate	High	High	Very High
AdjTN & AdjTP	Very Low	Low	Mixed	Mixed	Mixed	High	Very High
		excl 1				excl 7	
AdjTurb median	1.1	4.0	3.7	8.3	17.2	31.6	160.3
AdjTP median	0.029	0.043	0.095	0.067	0.093	0.259	0.546
AdjTN median	1.57	1.77	4.05	2.13	2.31	4.14	4.65
% Obs. Chla >30ug/liter	0.2%	2.1%	1.3%	2.3%	2.4%	4.9%	30%*
Algal Bloom Odds	1 in 540	1 in 47	1 in 74	1 in 44	1 in 42	1 in 20.5	1 in 3.3
Chla range	0.5 - 6.6	0.5 - 12.7	0.5 – 10.7	0.5 – 15.3	0.5 – 20.4	1.4 – 24.4	2.4 – 64.2
median	1.5	1.8	1.4	2.0	3.1	4.6	20.2
Turbidity IQR	1.9-2.6	4.0-6.6	3.6-6.4	9.0-10.2	14.0-30.0	16.1-88.5	65.0-283.8
median	2.2	5.3	5.0	9.6	19.0	33.5	164.5
DOC IQR**	2.03-2.84	1.94-2.98	1.54-2.76	2.04-3.14	1.93-3.52	too few	too few
median	2.4	2.4	2.1	2.4	2.6	too few	too few
TP IQR	0.022-0.051	0.031-0.073	0.054-0.192	0.046-0.108	0.063-0.145	0.229-0.335	0.451-0.921
median	0.034	0.048	0.100	0.072	0.098	0.263	0.552
TN IQR	1.17-2.44	1.46-2.63	3.92-4.90	1.71-3.46	1.93-3.30	4.00-5.00	4.48-5.65
median	1.80	2.00	4.31	2.36	2.55	4.40	4.91
AdjTN:AdjTP median	48.8	39.0	46.1	33.1	25.1	15.6	8.5
count	524	1367	439	392	1249	82	20

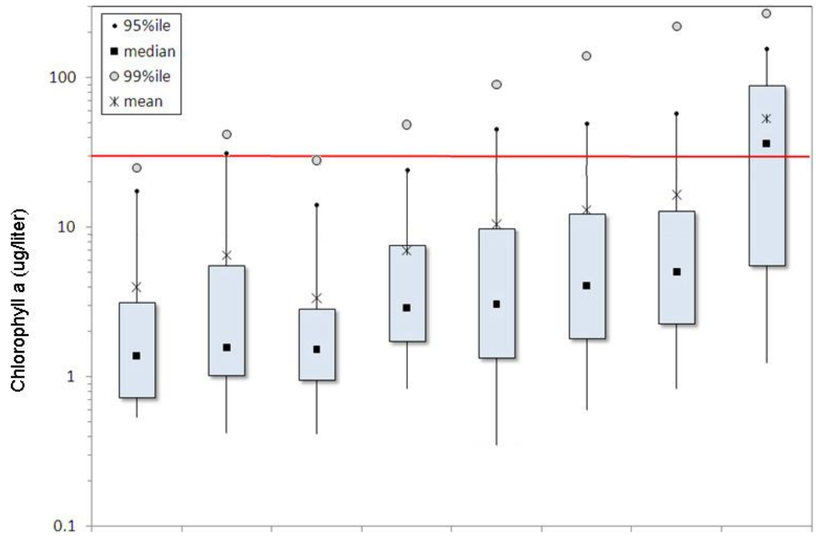
streams and rivers of all sizes. Extensive ditching to channel and drain surface flows has also significantly changed the historic stream flow patterns, and Strahler order seems to be a somewhat unreliable measure of branching complexity and stream size. The Strahler order requirement of order 5 and greater was dropped in the Coastal Plain and phytoplankton responses in the remaining analyses are for both streams and small rivers. The RPART models with the independent variables AdjTurb, AdjTP, AdjTN, and DOC were rerun on separate MACP and SEP data sets.

Figures 9 and **10** show the binning results for MACP and SEP streams and rivers, respectively. A gradient in the Chla response to the distinct bins is found in each bioregion. Bin 1 supports low median Chla concentrations of about 1.35 µg/liter. Bins 2 and 3 with low turbidity, low DOC, and nutrients levels slightly higher than those in Bin 1 also support low Chla concentrations. Medians ranged from 1.4 – 2.0 µg/liter, and none of the 380 SEP samples and just 5 of 322 (1.55%) MACP samples exceeded 30 µg/liter. Starting with bin 4, Chla medians and algal bloom frequencies increase steadily. Bin 7 (High) has a median Chla concentration of 4.9 µg/liter in MACP and 9.7 µg/liter in SEP. Algal bloom rates in bin 7 are 9.8% in MACP and 19% in SEP. A bin 8 in the MACP, with very high levels of nitrogen, phosphorus, turbidity, and DOC, could be assembled from available data. It had an n size of 21, a median of 37.3 µg/liter, and an algal bloom rate of 52.8%.

An effect of high DOC concentrations on the Chla nutrient response can be seen when the MACP intermediate bins in **Figure 9** are compared. Bins with similar AdjTurb levels but higher DOC concentrations have higher Chla concentrations. Specifically, bin 5 is higher than bin 3 and bin 6 is higher than bin 4. DOC concentrations are typically lower in SEP (**Appendix A Table 4**) and the DOC effect is not as pronounced. Still, the upper percentiles of Chla distributions are higher in bin 4 relative to bin 3 and in bin 6 relative to bin 5 (**Figure 10**).

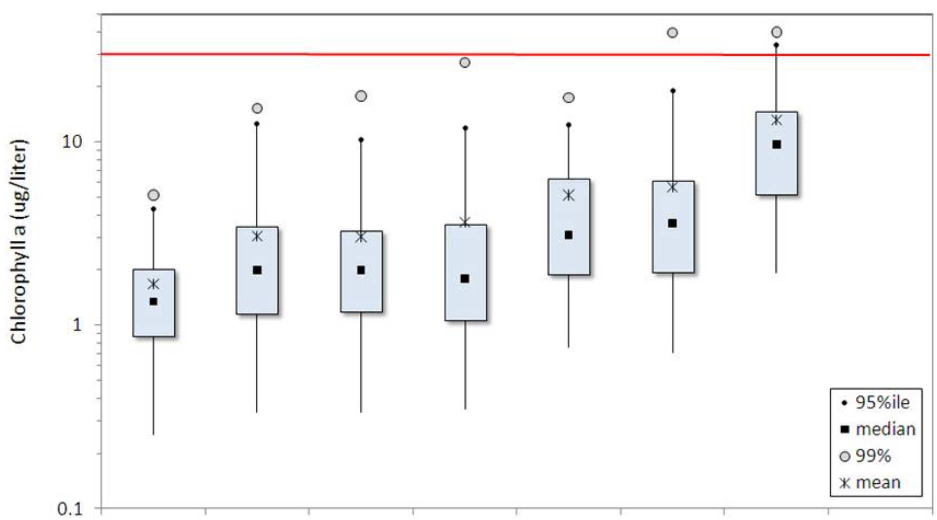
To further examine the suspected influence of DOC on the Chla nutrient response, the analysis data sets in the MACP and SEP were each divided into two DOC groups: <2.9 mg/liter and >2.9 mg/liter. (The MACP threshold of 2.9 mg DOC/liter was chosen for demonstration purposes but similar results occur with the SEP threshold of 3.4 mg DOC/liter.) These groups were then re-binned using the same AdjTurb, AdjTP, and AdjTN breakpoints. In the MACP, bins in the high DOC group had higher overall Chla concentrations than corresponding bins in the low DOC group (**Figure 11)**. In the SEP, the high DOC group showed a tendency towards higher values of the highest percentiles (e.g. 95th and 99th percentile) but values in corresponding bins were not significantly different ($p<0.01$). **Tables 5** and **6** give for the MACP and SEP bins, respectively, the median value of the AdjTurb, AdjTN, and AdjTP; the median and interquartile range (IQR) of Chla concentrations; the frequency of algal blooms; and the median N:P ratio calculated from the AdjTN and AdjTP. Within the prescribed boundaries of the bins, AdjTurb levels are not significantly different in corresponding bins of the high and low DOC groups but tend to be slightly higher in the high DOC groups. In both regions, AdjTP is typically higher in the high DOC groups and AdjTN higher in the low DOC groups. This difference in the nitrogen and phosphorus concentrations in the environment results in some sharply different N:P ratios between the high and low DOC groups. DOC concentrations are clearly related in several ways to differences in nutrient concentrations and to phytoplankton responses in the abiotic condition gradients represented by the bins, but regional differences are also an important factor.

A subset of 745 records in the PRV river data, or 18%, had associated DOC concentrations. Overall, the PRV rivers have lower DOC levels than SEP and MACP rivers but can range above 5 mg/liter (**Table 3**), particularly in the Monocacy and Patapsco rivers (**Appendix A Table 2**). We decided to examine DOC effects in PRV river data and the subset divided into a low DOC group (n = 513) and a high DOC group



BIN	1	2	3	4	5	6	7	8	
AdjTurb	<4.52	<7.4	<7.4	>=7.4	<7.4	>=7.4	>=7.4	>15.41	NTU
AdjTP	<0.0285	<0.036	mixed	mixed	mixed	mixed	>= 0.036	>0.0528	mg/liter
AdjTN	<2.087	<2.80	mixed	mixed	mixed	mixed	>=2.80	>4.39	mg/liter
DOC	<2.7	<2.9	<2.9	<2.9	>=2.9	>=2.9	>=2.9	>3.1	mg/liter
		excl. 1					excl. 8		
n	17	25	144	133	661	536	214	53	

Figure 9. Mid-Atlantic Coastal Plain bioregion binning results. Thresholds used to create the abiotic condition gradient bins are given below the graph. Red line indicates 30 µg/liter Chla. See text for details.



BIN	1	2	3	4	5	6	7	
AdjTurb	<9.4	<23.5	<23.5	<23.5	>=23.5	>=23.5	>=23.5	NTU
AdjTP	<0.0244	<0.0553	mixed	mixed	mixed	mixed	>=0.0553	mg/liter
AdjTN	<1.197	<2.42	mixed	mixed	mixed	mixed	>=2.42	mg/liter
DOC	<2.7	<3.4	<3.4	>=3.4	<3.4	>=3.4	>=3.4	mg/liter
		excl 1				excl 7		
n	30	272	108	820	23	198	21	

Figure 10. Southeastern Plain bioregion binning results. Thresholds used to create the abiotic condition gradient bins are given below the graph. Red line indicates 30 µg/liter Chla. See text for details.

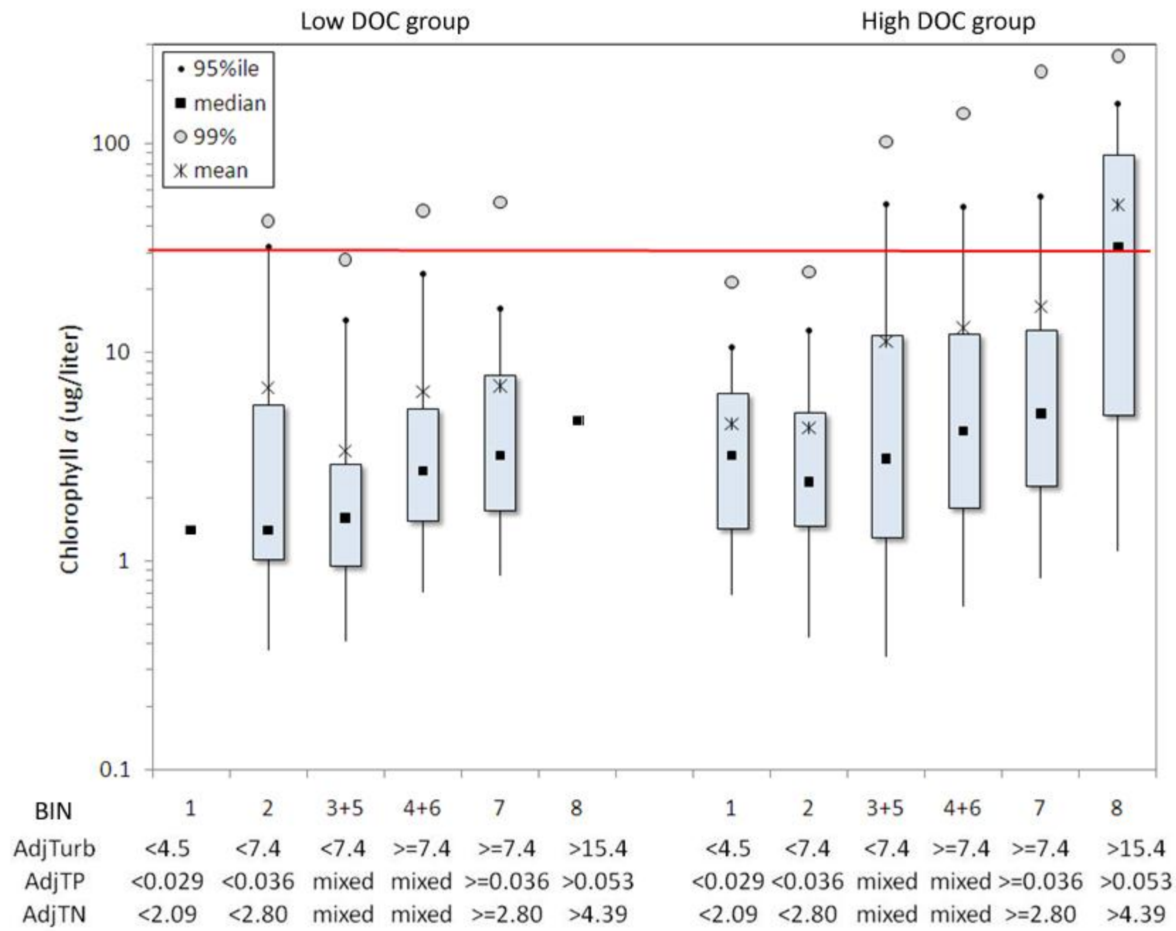


Figure 11. Mid-Atlantic Coastal Plain bioregion binning results for low DOC (<2.9 mg/liter) and high DOC (>2.9 mg/liter) groups. Thresholds used to create the abiotic condition gradient bins are given below the graph. Only medians are shown for bins with fewer than 20 samples. Red line indicates 30 µg/liter Chla. See text for details.

Table 5. Water chemistry in bins developed for streams and rivers in the Mid-Atlantic Coastal Plain bioregion, divided into high and low DOC groups. Median values for both adjusted and unadjusted turbidity, TP and TN are given for each bin. IQR: interquartile range (25th%ile – 75th%ile) of unadjusted (observed) turbidity, TP, and TN values. "Mixed" nutrient conditions have Hi P + Lo N and Lo P + Hi N. *, low sample size prevents representative calculation; **, two Chla concentrations > 30 ug/liter. Total n is 319 in the low group and 1446 in the high group. The bins correspond to those in **Figure 11**.

	Low DOC (<2.9 mg/liter)						High DOC (≥2.9 mg/liter)					
BIN	1	2	3+5	4+6	7	8	1	2	3+5	4+6	7	8
Description												
AdjTurb	Very Low	Low	Low	High	High	Very High	Very Low	Low	Low	High	High	Very High
AdjTN & AdjTP	Very Low	Low excl 1	Mixed	Mixed	High excl 8	Very High	Very Low	Low excl 1	Mixed	Mixed	High excl 8	Very High
AdjTurb median	1.9	4.7	2.9	12.5	13.5	22.1	1.9	4.6	3.9	15.5	13.9	24.3
AdjTP median	0.018	0.01	0.027	0.042	0.057	0.076	0.019	0.028	0.056	0.103	0.095	0.189
AdjTN median	1.47	2.15	5.13	2.6	4.44	5.46	1.37	1.94	3.09	1.55	3.6	6.53
% Chla >30ug/liter	*	8.7%**	1.4%	3.6%	3.2%	*	0.0%	1.9%	11.6%	11.2%	10.0%	50.0%
Algal Bloom Odds	*	*	1 in 72	1 in 27.5	1 in 31.5	*	*	1 in 53	1 in 8.6	1 in 8.9	1 in 10	1 in 2
Chla IQR	0.7 - 3.1	1.0 - 5.6	1.0 - 2.9	1.6 - 5.4	1.8 - 7.8	2.3 - 14.1	1.4 - 6.4	1.5 - 5.2	1.3 - 10.7	1.8 - 12.3	2.3 - 12.8	5 - 88.1
median	1.4	1.4	1.6	2.7	3.2	4.7	3.2	2.4	3.1	4.2	5.1	31.9
Turbidity IQR	1.9 - 3.5	3.0 - 6.0	2.0 - 5.0	10.0 - 17.5	11.0 - 17.5	17.5 - 29.7	1.3 - 3.3	3.0 - 5.8	2.6 - 5.7	11.0 - 23.9	9.6 - 23.9	20.0 - 33.6
median	2.0	5.0	3.1	12.6	13.7	22.4	2.0	4.7	4.0	15.6	14.1	24.8
DOC IQR	2 - 2.4	2.2 - 2.5	2 - 2.5	1.9 - 2.6	2 - 2.6	1.9 - 2.5	3.7 - 9.2	3.4 - 7.7	4 - 8.3	4.7 - 10.7	3.9 - 7.7	4.7 - 8
median	2.3	2.37	2.3	2.29	2.32	2.33	4.51	4.73	5.41	6.78	5.1	6.56
TP IQR	0.010 - 0.022	0.010 - 0.030	0.018 - 0.040	0.025 - 0.071	0.048 - 0.079	0.064 - 0.107	0.019 - 0.027	0.020 - 0.034	0.044 - 0.090	0.064 - 0.177	0.068 - 0.149	0.115 - 1.077
median	0.020	0.012	0.030	0.045	0.060	0.078	0.022	0.030	0.062	0.107	0.099	0.199
TN IQR	1.37 - 1.84	1.81 - 2.84	4.31 - 7.35	1.96 - 5.46	3.77 - 6.14	4.97 - 7.55	1.19 - 1.80	1.53 - 2.64	1.88 - 4.65	1.38 - 2.42	3.39 - 4.61	5.49 - 9.73
median	1.76	2.36	5.33	2.80	4.66	5.68	1.58	2.15	3.35	1.79	3.87	7.14
AdjTN:AdjTP median	79.29	173.22	231.72	50.85	78.5	80.27	64.42	81.05	47.25	13.74	36.83	28.14
count	19	23	144	55	63	15	29	53	579	536	211	56

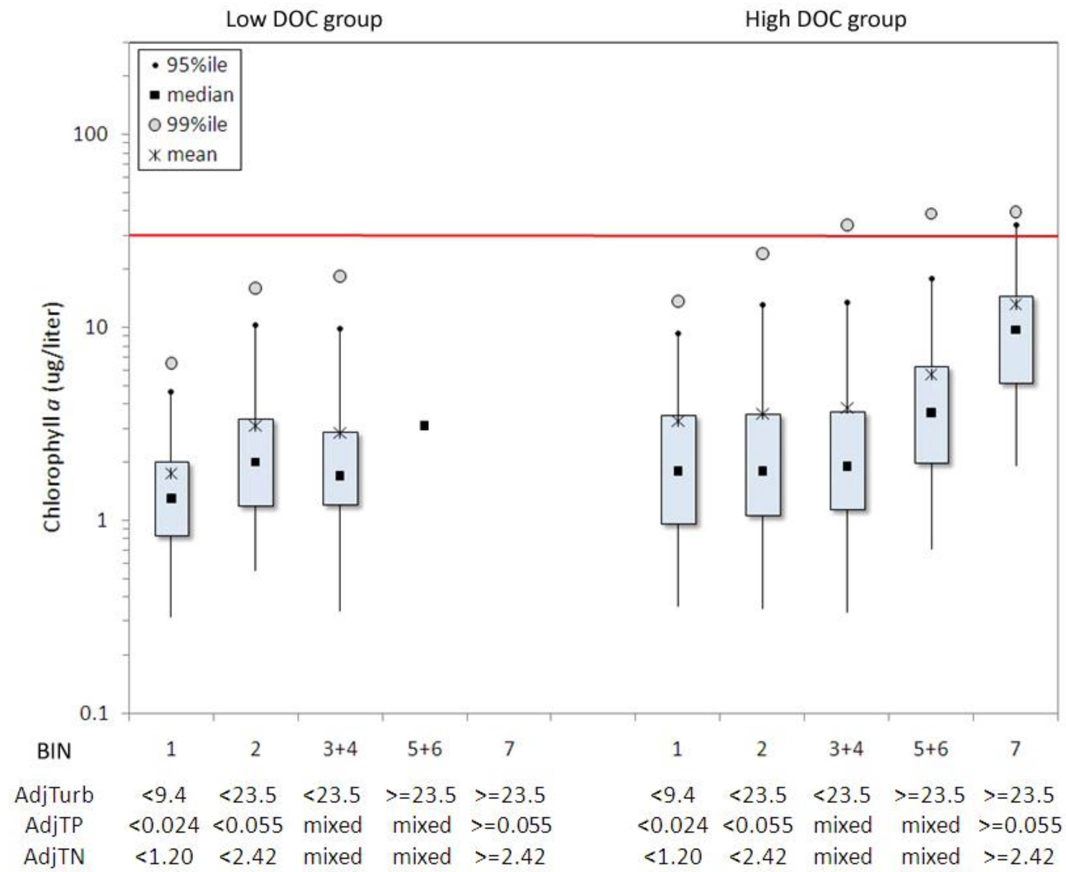


Figure 12. Southeastern Plain bioregion binning results for low DOC (<2.9 mg/liter) and high DOC (>2.9 mg/liter) groups. Thresholds used to create the abiotic condition gradient bins are given below the graph. Red line indicates 30 µg/liter Chla. See text for details.

Table 6. Water chemistry in bins developed for streams and rivers in the Southeastern Plain bioregion divided into high and low DOC groups. Median values for both adjusted and unadjusted turbidity, TP and TN are given for each bin. IQR: interquartile range (25th%ile – 75th%ile) of unadjusted (observed) turbidity, TP, and TN values. "Mixed" nutrient conditions have Hi P + Lo N and Lo P + Hi N. *, low sample size prevents representative calculation. Total n is 270 in the low group and 1202 in the high group. The bins correspond to those in **Figure 12**.

	Low DOC (<2.9 mg/liter)					High DOC (≥ 2.9 mg/liter)				
BIN	1	2	3+4	5+6	7	1	2	3+4	5+6	7
Description AdjTurb AdjTN & AdjTP	Very Low Very Low	Low Low excl 1	Low Mixed	High Mixed	High High	Very Low Very Low	Low Low excl 1	Low Mixed	High Mixed	High High
AdjTurb median	3.5	5.7	6.1	31.8		4.6	8	8.8	40.7	133.8
AdjTP median	0.013	0.021	0.057	0.047		0.017	0.032	0.083	0.086	0.202
AdjTN median	0.77	1.71	2.45	1.14		0.63	0.78	0.94	1.03	2.87
% Chla >30ug/liter	0.0%	0.0%	0.0%	*		0.6%	0.6%	1.3%	1.5%	15.8%
Algal Bloom Odds	-	-	-	*		1 in 166	1 in 164	1 in 76	1 in 67	1 in 6.3
Chla IQR	0.8 – 2.0	1.2 - 3.4	1.2 - 2.9	1.7 - 4.8		1.0 - 3.4	1.1 - 3.6	1.1 - 3.7	1.9 - 6.3	5.0 - 14.4
median	1.3	2.0	1.7	3.1		1.8	1.8	1.9	3.6	9.7
Turbidity IQR	0.3 – 8.0	1.1 - 16.7	1.0 - 16.5	24.2 - 68.3		0.9 - 8.4	1.6 - 19.2	1.3 - 22	25.0 - 129.6	24.1 - 473.2
median	3.6	5.9	6.25	31.95		4.7	8.15	8.9	40.8	134
DOC IQR	1.49 - 2.87	1.89 - 2.84	1.62 - 2.84	2.13 - 2.64		3.00 - 6.98	3.06 - 8.99	3.20 - 10.29	3.44 - 11.29	3.54 - 12.94
median	2.55	2.44	2.37	2.44		4.05	4.55	4.85	5.88	6.29
TP IQR	0.009 - 0.025	0.012 - 0.056	0.012 - 0.137	0.025 - 0.102		0.011 - 0.026	0.016 - 0.055	0.053 - 0.28	0.029 - 0.279	0.072 - 0.625
median	0.016	0.023	0.059	0.052		0.02	0.035	0.085	0.088	0.205
TN IQR	0.45 - 1.36	0.47 - 2.56	0.49 - 4.40	0.45 - 2.12		0.40 - 1.37	0.42 - 2.14	0.39 - 3.70	0.66 - 2.35	2.68 - 4.23
median	0.99	1.93	2.67	1.36		0.84	1.00	1.19	1.27	3.24
AdjTN:AdjTP median	60.11	81.79	45.48	20.74		37.30	23.10	10.42	10.87	15.21
count	41	136	79	14	no data	168	496	310	207	21

Table 7. Water chemistry in the river bins of the Piedmont, Ridges, and Valleys region, for 745 records with associated dissolved organic carbon data. Median values for both adjusted and unadjusted turbidity, TP and TN are given for each bin. IQR: interquartile range (25th%ile – 75th%ile) of unadjusted (observed) turbidity, TP, and TN values. "Mixed" nutrient conditions have Hi P + Lo N and Lo P + Hi N. *, low sample size prevents representative calculation. Total n is 233 in the low group and 513 in the high group. Bins 6 and 7 were combined due to little or no data.

	Low DOC (<2.9 mg/liter)						High DOC (≥2.9 mg/liter)					
	1	2	3	4	5	6+7	1	2	3	4	5	6+7
Description AdjTurb	Very Low	Low	Low	Moderate	High	High & VHigh	Very Low	Low	Low	Moderate	High	High & VHigh
AdjTN & AdjTP	Very Low	Low excl. 1	Mixed	Mixed	Mixed	High & VHigh	Very Low	Low excl. 1	Mixed	Mixed	Mixed	High & VHigh
AdjTurb median	1.0	3.9	3.5	8.7	15.5		0.9	3.9	4.3	8.3	16.9	635
AdjTP median	0.010	0.014	0.028	0.031	0.041		0.026	0.041	0.084	0.082	0.091	0.761
AdjTN median	2.06	2.12	3.94	2.20	2.52		1.37	1.53	3.60	2.13	2.50	4.24
% Chla >30ug/liter	0.0%	0.0%	0.0%	0.0%	0.0%		0.0%	1.5%	0.0%	0.0%	1.0%	*
Algal Bloom Odds	-	-	-	-	-		*	1 in 68	*	*	1 in 96	*
Chla IQR	1.4 - 2.9	1.3 - 3.4	0.8 - 2.2	1.6 - 3.2	1.7 - 4.2		1.1 - 2.2	1.4 – 4.0	*	1.6 - 3.8	1.6 - 3.9	9.7 – 13.9
median	2.1	2.0	1.3	2.2	2.4		1.6	2.3	1.3	2.6	2.8	12.8
Turbidity IQR	1.0 - 2.8	4.0 - 6.3	3.0 - 6.0	9.0 - 10.0	13.0 - 24		2.0 - 2.0	4.0 - 7.0	4.4 - 6.3	9.0 - 10.0	13.0 - 30.3	*
median	2.0	5.0	4.65	10.0	17.0		2.0	5.0	5.5	10.0	18.0	637
DOC IQR	1.83 - 2.51	1.83 - 2.45	1.29 - 2.49	1.76 - 2.43	1.73 - 2.59		3.62 - 4.97	3.15 - 4.28	3.12 - 3.62	3.16 - 5.47	3.41 - 4.77	*
median	2.24	2.08	1.82	2.16	2.15		4.02	3.73	3.32	4.11	3.81	7.09
TP IQR	0.009 – 0.025	0.013 – 0.030	0.025 – 0.044	0.021 – 0.048	0.033 – 0.070		0.019 – 0.04	0.031 – 0.076	0.055 – 0.251	0.048 – 0.15	0.062 – 0.129	*
median	0.015	0.019	0.033	0.036	0.046		0.031	0.046	0.089	0.087	0.096	0.766
TN IQR	0.97 - 2.61	0.97 - 2.8	3.95 - 4.7	1.29 - 3.66	1.45 - 4.06		0.73 - 2.47	0.9 - 3	1.67 - 3.91	1.05 - 2.64	2.34 - 3.3	*
median	2.30	2.35	4.20	2.44	2.76		1.59	1.76	3.85	2.37	2.74	4.5
AdjTN:AdjTP median	139.3	96.4	164.6	67.2	53.9		65.4	36.7	43.6	21.6	31.5	4.0
count	82	170	46	51	159	no data	26	68	12	26	96	5

(n = 233) at 2.9 mg/liter. The results are summarized in **Table 7**. In the PRV rivers, AdjTN is again higher in the low DOC group and AdjTP is typically higher in the high DOC group, leading to distinct differences in the N:P ratios. Median Chla values in corresponding bins are not different but—like SEP—the upper quartiles tend to be higher in the high DOC group.

For sampling stations in the coastal plain bioregions, it is not possible to develop the relationship between the frequency of bins with high nutrients and the frequency of algal blooms shown in **Figure 8** for PRV river stations. Most of the sampling locations in this region are randomly selected and fixed sampling locations are relatively uncommon. The bins to project how often algal blooms will occur in a given stream size under certain environmental nutrient and light conditions.

4.3.3 Adjusted and unadjusted concentrations of turbidity, TN, and TP in bins

Up to this point in the analysis, the bins forming the abiotic condition gradient have been characterized by their chlorophyll-adjusted values. These adjusted values more accurate represent environmental conditions surrounding the phytoplankton cells because the phytoplankton components of turbidity, TP, and TN are minimized. Characterizing streams and rivers by their bin frequencies should continue to rely on adjusted values to avoid misrepresenting trophic state and implementing inappropriate management strategies. More sophisticated methods than the one used in this study could be employed to chlorophyll-adjust the parameters. However, in practical applications adjustments would not be necessary when chlorophyll *a* concentrations are low. There are several possible methods for converting AdjTurb, AdjTP, and AdjTN values back to unadjusted values. Converting the values is necessary in order to compare boundaries of the abiotic condition gradient bins for phytoplankton to those for periphyton and macroinvertebrates. We chose a direct method and calculated the median and interquartile (25^{th} – 75^{th} percentile) of the observed (unadjusted) turbidity, TP, and TN values in each bin of the analysis data set. The unadjusted median and range values are listed in **Tables 4 – 6**.

4.4 Findings

A series of water quality "bins" with distinct nutrient and water clarity conditions was created for each physiographic region using breakpoints generated in recursive partitioning model runs. The bins have chlorophyll *a* concentrations that grade toward degradation as the non-algal (adjusted) constituents of phosphorus, nitrogen, and turbidity (and dissolved organic carbon in the coastal plain) increase. The pattern of rising Chla concentrations is consistent across physiographic regions despite differences in gradient, retention time, nutrient sources, and coloring of the water by organic compounds (**Figures 7, 9-12**). A relationship between median Chla and frequency of algal blooms is found in each region.

The frequency of algal blooms (Chla >30 µg/liter) climbs quickly when nutrient concentrations exceed about 0.041 – 0.103 mg/liter AdjTP *and* 1.03 – 2.60 mg/liter AdjTN *and* water clarity is affected by relatively high levels of non-algal turbidity and/or color. When nutrient concentrations are both below these thresholds *and* water clarity is good, blooms are very infrequent and median Chla concentrations range between 1.3 and 3.2 µg/liter. The binning results demonstrate the importance of examining phytoplankton responses to nitrogen and phosphorus in the context of the light environment. When non-algal turbidity and coloring levels are low, making the water column relatively transparent, and nitrogen and/or phosphorus exceed the thresholds above, Chla remains low. This Low Turbidity-Mixed Nutrients case is represented by bins 3 and 4 in the PRV rivers, bins 3+4 in the SEP ("western shore") streams and rivers, and bins 3+5 in the MACP ("eastern shore") streams and rivers. The distribution of Chla concentrations rises noticeably when non-algal turbidity and/or coloring are high, even if one or both nutrients are below the thresholds and could be controlling (limiting) algal growth. This latter case, called the High Turbidity-Mixed Nutrients bin, is represented by bin 5 in PRV, bin 5+6 in SEP, and bin 4+6 in MACP. Median Chla concentrations in the bins range between 2.4 and 4.2 µg/liter. Once AdjTP and AdjTN both exceed the thresholds by a substantial margin (e.g., 1.5x), the Chla distribution rises sharply regardless of any light limitation of photosynthesis caused by the high turbidity or coloring.

The coincidence of high Chla with high turbidity in the High and Very High bins (**Tables 4-6**) raises the question – are algal concentrations elevated because they were scoured from the periphyton or washed in from upstream lakes and reservoirs during storms, or are they real responses to stream and river conditions? Turbidity and nutrient spikes in streams and rivers are expected after storm events, and both the rareness of High and Very High bins and the occasionally extreme turbidity levels in these bins tend to support the storm interpretation. However, non-algal nutrient concentrations in these bins strongly overlap those in the High Turbidity-Mixed Nutrients bin of each region. Furthermore, the High Turbidity-Mixed Nutrients bin is the second most frequent bin in PRV (30.7%) and MACP (33.1%) and is common in SEP (**Table 8**). Some of the elevated Chla concentrations in the high turbidity bins are undoubtedly related to storm events, but because the combination of high turbidity and relatively high nutrients is found often, the condition is not necessarily related to storm events and eutrophication is also a probable cause of the elevated Chla.

The overall frequency of algal blooms found in each physiographic region appears to be related to the proportion of high turbidity/high nutrient(s) bins in the region. As shown in **Table 8**, just 16.4% of SEP rivers and streams samples fall in the bioregion's high turbidity/high nutrient(s) bins (i.e., 5, 6, and 7), and the bioregion has the lowest overall frequency of algal blooms (1.02%), or about 1 in every 100 samples. A much higher percent, 52.5%, of samples are classified in the high turbidity/high nutrient(s) bins of MACP streams and rivers, and overall algal bloom frequency in this bioregion is comparatively high at 10.38%, or about 1 in every 10 samples. Looking only at rivers in the PRV region, 33.2% of samples fall in the region's high turbidity/high nutrient(s) bins and the overall algal bloom frequency is 2.07%. A heightened Chla response to an increasing frequency of the high turbidity/high nutrient(s) bins is also seen on a station basis in the PRV, where long-term monitoring at fixed stations has been performed. As the frequency of bins 5-7 experienced at a station increases, the frequency of algal blooms increases (**Figure 8**). This relationship between the frequency of algal blooms and the frequency of high turbidity/high nutrient(s) bins in a river segment could be useful in developing river management scenarios.

In addition to the scattering and absorption effects of non-algal turbidity on phytoplankton light environments, AdjTurb is--not surprisingly--related to AdjTP concentration in all physiographic regions ($p<<0.01$). The sorption of

Table 8. Frequency of bins representing the abiotic condition gradient in the Piedmont, Ridges, and Valleys (PRV), the Southeastern Plain (SEP), and the Mid-Atlantic Coastal Plain (MACP).

Description							
AdjTurb	VLow	Low	Low	Moderate	High	High	VHigh
Adj N + P	VLow	Low	Mixed	Mixed	Mixed	High	VHigh
PRV rivers							
bin #	1	2	3	4	5	6	7
all	12.9%	33.6%	10.8%	9.6%	30.7%	2.0%	0.5%
SEP streams and rivers							
bin #	1	2	3+4		5+6	7	
all	14.2%	42.9%	26.4%		15.0%	1.4%	
high DOC	14.0%	41.3%	25.8%		17.2%	1.7%	
low DOC	15.2%	50.4%	29.3%		5.2%		
MACP streams and rivers							
bin #	1	2	3+5		4+6	7	8
all	2.7%	4.3%	40.5%		33.1%	15.4%	4.0%
high DOC	2.0%	3.6%	39.5%		36.6%	14.4%	3.8%
low DOC	6.0%	7.2%	45.1%		17.2%	19.7%	4.7%

phosphorus on suspended particles is a well-recognized phenomenon regulated chiefly by particle size and composition, pH, and the surrounding ionic concentrations. Extremely high AdjTP rarely occurs with extremely low AdjTurb and *vice versa*. It is therefore difficult in natural, uncontrolled systems to separate the effect of increasing turbidity on the phytoplankton light environments from a phytoplankton response to the total phosphorus concentrations normally associated with turbidity. We posit that the vertical light gradient created by particles and dissolved substances in the water column is an important controller of phytoplankton populations in rivers because it is in lakes, estuaries and oceans. The binning process keeps the two parameters together, grouping low turbidity with low TP and high turbidity with high TP, and as such resembles natural systems.

The analysis demonstrated a strong DOC effect on the Chla-nutrient-turbidity interactions in the MACP bioregion and a generally muted DOC effect on interactions in the SEP and PRV regions. The MACP bioregion has by far the largest percentage of high DOC streams, followed by the SEP bioregion and then the PRV region. We initially hypothesized a significant DOC effect on phytoplankton responses to nutrients and light would occur through DOC's effect on light. Dissolved organic substances shift the underwater light spectrum to shorter wavelengths, coloring the water brown and attenuating light in the range required for photosynthesis (400-700 nm). High DOC is also indicative of organic compounds that tend to be more refractory in nature and thus more difficult for bacteria and fungi to metabolize. Aquatic systems with naturally high inputs of organic compounds from fringing wetlands and marshes as well as forest leaf litter typically have a high proportion of humic acids, which are low in nitrogen content. Furthermore, DOC tends to slow the sedimentation rates of water column phosphorus through stabilization of colloidal particles. The net result can be waters with low carbon-to-nitrogen and N:P ratios. The comparatively low N:P ratios in each of the three regions' High DOC group indicates at the very least some action of DOC on nitrogen and phosphorus stoichiometries. In the MACP, with its particularly high N:P ratios and overall higher nitrogen concentrations, a relatively large source of nitrogen is indicated.

Conductivity is another environmental factor that may be influencing TP in the MACP bioregion. Conductivity in streams and small rivers is generally lower in MACP compared to SEP (**Appendix A Tables 3** and **4**). The divalent cations of calcium and magnesium, two constituents of conductivity, are known to react with ortho-phosphate, the soluble form of phosphorus taken up by algal cells, and form a precipitate. The precipitate tends to settle out of the water column. In waters with low conductivity, fewer calcium and magnesium cations are present to interact with ortho-phosphate, thus proportionally more of the TP in the water column is composed of ortho-phosphate. In these cases, a lower concentration of TP is adequate for a comparable level of ortho-phosphate uptake by algae. Note that in the low AdjTurb-mixed nutrients bins, TP concentrations in MACP (bin 3+5) are lower than those in the equivalent SEP conditions (bin 3+4). Since bins higher than these experience blooms and bins lower than these do not, we postulate that nutrient conditions in these bins are at or near the thresholds which limit algal bloom formation. The lower medians of TP concentration in the MACP bins 3+5 would suggest they have a higher fraction of ortho-phosphate than the SEP bins 3+4. The fractions were not examined in this study to confirm this hypothesis, but the analysis would be a worth-while issue to investigate as it may explain why algal blooms occur lower TP concentrations in the MACP.

5. Periphyton

Eutrophication can be expressed as excess growth of periphyton in streams and rivers. Periphyton, or biofilm, is a general term used to describe the microbial community of algae, bacteria, and fungi growing on nearly all submerged surfaces, including sediment grains, rocks, detritus, vascular plants, tree roots, and even animals. Periphyton support a broad size range of grazers, from protozoans and macroinvertebrates to snails and fish. Grazers are important in keeping periphyton biomass in check. The algae and some bacteria of the periphyton are autotrophs and require well-lit or shallow habitats in order to photosynthesize; the other constituents are heterotrophs and can live in heavily shaded or deep habitats. Heterotrophic periphyton are primarily responsible for decomposing heavy organic inputs from the tree canopy (leaf litter) in forested headwater streams and from fringing marshes and wetlands in larger streams and rivers.

Periphyton are a possible indicator of nutrient pollution because their growth is stimulated by the addition of inorganic and organic nutrients. If periphyton metrics calculated from the monitoring data can be linked to corresponding benthic macroinvertebrate metrics, it may be possible to demonstrate with monitoring data a connection through the food web between nutrients and aquatic life use impairment. If such a link is demonstrated, one question to ask is “can a nutrient threshold be identified above which periphyton production is too great for the grazer community to control?”

5.1 Periphyton data provided to ICPRB

Maryland and neighboring states collect a range of periphyton data. The data include algal mat biomass, biovolume, and nutrient content; diatom taxa identifications, counts, and metrics; and habitat parameters, including available light. In this study, we were provided periphyton and associate data from six sources for all years available. Sampling locations where these data were collected are identified in **Figure 13**.

Periphyton require light and nutrients for growth, and therefore we focused our analysis on the relationship between the biomass of the bottom algal mat, light or some surrogate thereof, and total nitrogen and total phosphorus. Parameters collected by each data source vary depending on the intended use of the data. Diatom taxa data were collected in Maryland. USGS collected periphyton data, but no information on shading. Virginia, Delaware, Pennsylvania, and SRBC each collected a variety of information on periphyton, water quality, and habitat.

VADEQ made periphyton “bulk” measurements, namely ash free dry mass (AFDM), chlorophyll *a* content, and nitrogen and phosphorus content of the bottom algal mat. Water quality data, including nutrients, were also collected and while light was not directly measured, habitat parameters that could serve as surrogates for light such as bank and riparian vegetation were scored. There are many locations where each of these parameters was collected but no location where all of these parameters were collected synchronously.

Delaware collected the biovolume, taxonomic counts, and cell density of the bottom algal mat, shading, and water column parameters that included total nitrogen, total phosphorus, and turbidity. There were many stations where each of these parameters was collected, but no location where all parameters were collected on the same date. Periphyton data were typically collected in May and June, whereas habitat data were collected in October and November.

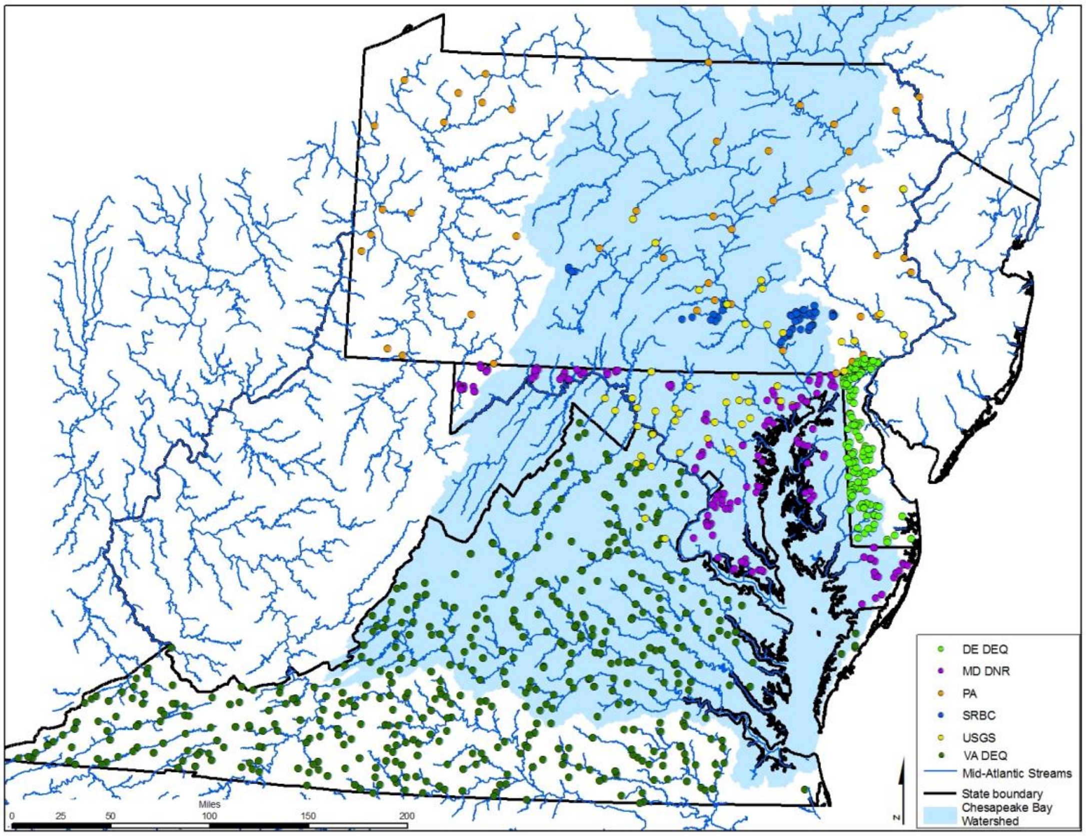


Figure 13. Periphyton sampling locations identified in the available data sets.

Pennsylvania collected taxa counts, biovolume, abundance, and the chlorophyll *a* and nutrient content of the bottom algal mat. Pennsylvania also provided data on percent canopy cover, and total nitrogen and total phosphorus. There were 47 occurrences of all three of these parameters being collected at the same location and time and approximately 150 occurrences of all three parameters collected at the same location but at different sampling times.

The Susquehanna River Basin Commission (SRBC) collected information on total volume, chlorophyll *a* content, carbon content, and nutrient content of the bottom algal mat, and provided data on a variety of habitat measurements, including percent canopy cover and bank vegetation, as well as total nitrogen and total phosphorus. Unfortunately the station identification provided with the periphyton data could not be linked to the water quality and habitat data that resided in another database.

5.2 Analysis of selected Virginia periphyton data

Virginia had by far the most extensive coverage of periphyton sites but nutrient samples were collected at different times and/or places. To explore the data for possible nutrient responses, the requirement of synchronous biological, habitat, and water quality sampling at the same site applied elsewhere in this study was relaxed.

With the exception of three samples collected in August, periphyton were typically sampled during September through November in the years 2004 through 2008. Habitat data were collected at each site from April through June and in October and November during the same years. We assumed the habitat parameters (i.e., bank stability, bank vegetation, riparian vegetation, available cover, and stream channel alteration) changed little during a given year, and therefore habitat and periphyton data were matched by sample location and year.

Because periphyton growth can respond to nutrients throughout the year, it was felt that some average nutrient value might adequately represent nutrient conditions and allow periphyton samples to be matched with water quality data. Each periphyton sampling site was matched to all water quality stations located on the same stream reach within a five mile radius, within the same ecoregion and stream order. Water quality data collected up to one year prior to the periphyton collection date were averaged. TN:TP ratios at each site were calculated before the data were averaged. Although it would have been meaningful to include DOC and turbidity data, not enough of these data were available to establish meaningful relationships with the algal biomass.

Periphyton and/or water quality data collected in tidal or impounded (lakes, dams) waterbodies were excluded from the analysis. The decision for exclusion was based on station descriptions and GIS analysis. Ecoregion and Strahler stream order were assigned to each station using GIS analysis and the 2005 National Hydrography Dataset Plus (NHDPlus) layer http://www.horizon-systems.com/nhdplus.

Values indicated as below detection limit BDL were taken as recorded. A single outlier AFDM data point (1,162 g/m^2) was removed. Data records were also removed from the analysis data set if they under-represented their bioregions and stream orders. Specifically, 14 records in the coastal plain bioregions and 13 records for Strahler orders 5 and 6 were removed. The remaining 139 records were distributed across both bioregion and Strahler order, and were the basis for all of the following analyses (**Table 9**). Support for the decision to remove these data was found in several exploratory analyses which split out the two coastal plains and the two largest stream orders.

5.2.1 Exploratory analysis

Periphyton chlorophyll content (CHL_BEN), phosphorus content (TP_ALG), and AFDM in the selected data set was individually matched to an array of independent variables and run through an RPART model for the purpose of detecting and minimizing important confounding factors. The independent variables are listed in **Table 10.** As confounding factors were identified, the original data set was iteratively filtered until nutrients became the primary variables splitting the data. Minimum size of the split nodes was variably set to 20 or 30 to observe how this user-imposed "rule" affected model results.

Benthic chlorophyll (mg/m^2)

Several RPART models were constructed with the available variables to investigate possible nutrient breakpoints for benthic chlorophyll. One RPART model run was done initially on the separate Piedmont, Ridges, and Valleys data sets. It pointed to stream bank related parameters as important splitting factors, with splits occurring in the middle of the 0-20 range. In the Virginia habitat rating system adapted from Plafkin *et al.* (1989) Rapid Bioassessment Protocol, habitat scores at

Table 9. Data records selected for periphyton analysis. Highlighted records under-represent their bioregion and stream order, and were excluded from the data analysis. MACP, Mid-Atlantic Coastal Plain; SEP, Southeastern Plain.

Strahler order	Bioregion				
	MACP	SEP	Ridges	Piedmont	Valleys
1	1	3	12	17	4
2		4	10	22	8
3		2	8	14	10
4		4	4	19	11
5			2	6	2
6				2	1

Table 10. List of independent variables used in the exploratory RPART model runs. * metric is scored on a 0-20 point scale. Habitat assessment protocols are adapted from Plafkin *et al.* 1989.

Parameter	Description
Region	Bioregion (Note: sites in the Valleys bioregion are separated into with and without karst geology groups)
SS	Strahler order
STREAMSZ	Strahler order groups: SM (small), 1-3; M (medium), 4; LG (large), 5-6
pH_1y	1 year average of pH measurements in neighboring stream locations
COND_1y	1 year average of conductivity measurements in neighboring stream locations
TN_1y	1 year average of total nitrogen concentrations in neighboring stream locations
TP_1y	1 year average of total phosphorus concentrations in neighboring stream locations
TNTP_1y	Ratio of TN1y to TP1y
FLOW	Wetted width of stream and substrate exposure*
ALTER	Large scale changes in shape of the stream channel*
BANKS	Bank erosion, and the potential for erosion*
BANKVEG	Amount of vegetative protection on stream bank and on near-stream portion of riparian zone*
COVER	Epifaunal substrate and available cover provided by fallen trees, banks, boulders, etc.*
SED	Amount of sediment that has accumulated in pools*
RIPVEG	Width of the natural vegetation from edge of the stream bank through the riparian zone*
EMBED	Extent to which hard substrate is covered by silt, sand and mud*
TotHabSc	VADEQ habitat score derived from 10 parameters (see for example Burton and Gerritsen, 2003, for details)

this level are at or near the lower boundary of sub-optimal. (An optimal score for the habitat metrics is 16-20, sub-optimal is 11-15, marginal is 6-10, and poor is 0-5.) To remove the various bank influences, the data set was filtered to exclude all records with BANKS < 11, BANKVEG < 11, and RIPVEG < 11. The three variables were identified in the RPART trees as important factors. This step coincidently removed records with ALTER < 11. The filtered Piedmont, Ridges, and Valleys data were combined and the model rerun. It strongly separated karst and non-karst sites. Samples associate with karst (n=11) were removed and the model rerun again. The first split of this model run was on TN at 0.47 mg/liter, but it also identified another ("next best") TN split at 0.415 mg/liter and TP splits at 0.0475 and 0.0558 mg/liter. CHL_BEN was significantly higher in the samples with higher TN ($p<0.01$).

In another RPART model, Valley was excluded from the analysis on the justification that its higher pH and conductivity levels distinguish it from the Piedmont and Ridges bioregions. The model runs again pointed to bank-related features as important split parameters, and the Piedmont and Ridges data were filtered so that all remaining records had ALTER, BANKS, BANKVEG, RIPVEG, and COVER scores $\geq$10 (optimal and suboptimal quality). The first split in this model run was on TP at 0.0425 mg/liter and the second split was on TN at 0.525 mg/liter. "Next best" splits identified on various tree branches were 0.575 and 0.195 mg/liter for TN and 0.025 mg/liter for TP.

Additional RPART model runs with different combinations of independent factors, and generally support the finding that when stream bank conditions are optimal or suboptimal and sampling sites are not located on karst geology, Piedmont, Ridges, and Valleys streams had breakpoints in the range of 0.0425 - 0.0558 mg/liter for TP and 0.415 – 0.575 mg/liter for TN, and another set of lower breakpoints.

Algal mat phosphorus content (mg/liter)

The algal mat phosphorus content, TP_ALG, showed a very clear response to water column TP regardless of habitat condition or bioregion. TP_1y was the first and often the second split in various RPART model runs, with TN_1y frequently identified as the second split in the trees or a "next best" split in the RPART summaries. Breakpoints for TP_1y ranged from 0.0475 - 0.0658 mg/liter, with some low breakpoints at ~0.02 mg/liter. Breakpoints for TN_1y ranged widely from 0.185 - 0.915 mg/liter. A highly significant log-log regression between water column TP_1y and TP_ALG was found in these streams (r^2 0.48, p<<0.001).

Ash-free dry mass (g/m^2)

Exploratory RPART model runs with the response variable ash-free dry mass of the periphyton, or AFDM, did not develop like those for CHL_BEN or TP_ALG. No clear branching tendencies were found in the first splits. Total habitat score and pH were most often the first split parameters, followed by other habitat parameters, Strahler order, and occasionally TN. A slight decrease in AFDM occurs with improving habitat condition, expressed as Virginia's total habitat score, and a slight increase occurs with increasing Strahler stream order. No relationship is found with pH. Overall, it did not appear that the nutrient TN or TP played an important role in AFDM.

5.2.2 Nutrient bins and results

A total of five distinct nutrient bins were successfully created using thresholds that approximate the RPART nutrient breakpoints. The bin thresholds are given in **Table 11**. The bins were applied to the entire data selected for analysis ("unfiltered") as well as to a subset of the data that had been thinned of records that the RPART trees indicated may be significantly affected by non-nutrient parameters. Nutrient responses in the removed records were thought to be confounded by those other parameters. The unfiltered data set has all Virginia records for stream Strahler order 1-4 in the Piedmont, Ridges, and Valleys bioregions. Records in remaining "filtered" data set have habitat scores >10 and no karst geology at the sampling location. **Figure 14** shows the binning results for CHL_BEN, TP_ALG, and AFDM for both the unfiltered and filtered data. **Table 12** presents the median values of TN_1y, TP_1y, TN:TP_1y, COND_1y and Virginia's total habitat score for each bin.

CHL_BEN and TP_ALG showed similar, increasing responses to increasing nutrients in the filtered data set. Bin 1 with the lowest concentrations of both TN and TP was associated with the lowest concentrations of the two periphyton parameters. Bin 5 with the highest TN and TP concentrations generally had higher values. TP_ALG was especially responsive to the nutrient gradient. Neither CHL_BEN nor TP_ALG seem to show any particular sensitivity to the contrasting bins 3 (high N:P) and 4 (low N:P), although CHL_BEN in bin 3 was somewhat higher than in bin 4 and even bin 5. The small sample sizes in most of the bins preclude a definitive result.

In the unfiltered data, CHL_BEN distributions are not as clear cut and variability in each bin is much greater. This indicates the confounding factors of karst and/or marginal or poor bank-related habitat parameters are interfering in some way with the CHL_BEN nutrient responses. The TP_ALG nutrient responses are apparently not affected by the confounding environmental factors. Changes in the bin

Table 11. Nutrient bins thresholds based on nutrient breakpoints observed in RPART trees. Number of observations per bin for CHL_BEN and TP_ALG are given for the filtered and unfiltered data sets.

		TP_1y	TN_1y	CHL_BEN count		TP_ALG count	
Description	BIN	(mg/liter)	(mg/liter)	filtered	unfiltered	filtered	unfiltered
Very Low	B1	< 0.025	< 0.25	20	24	20	24
Low	B2	< 0.04	< 0.43	13	29	13	28
LowTP, HighTN	B3	< 0.04	≥ 0.43	23	49	23	50
LowTN, HighTP	B4	≥ 0.04	< 0.43	7	9	7	9
High	B5	≥ 0.04	≥ 0.43	11	24	10	22

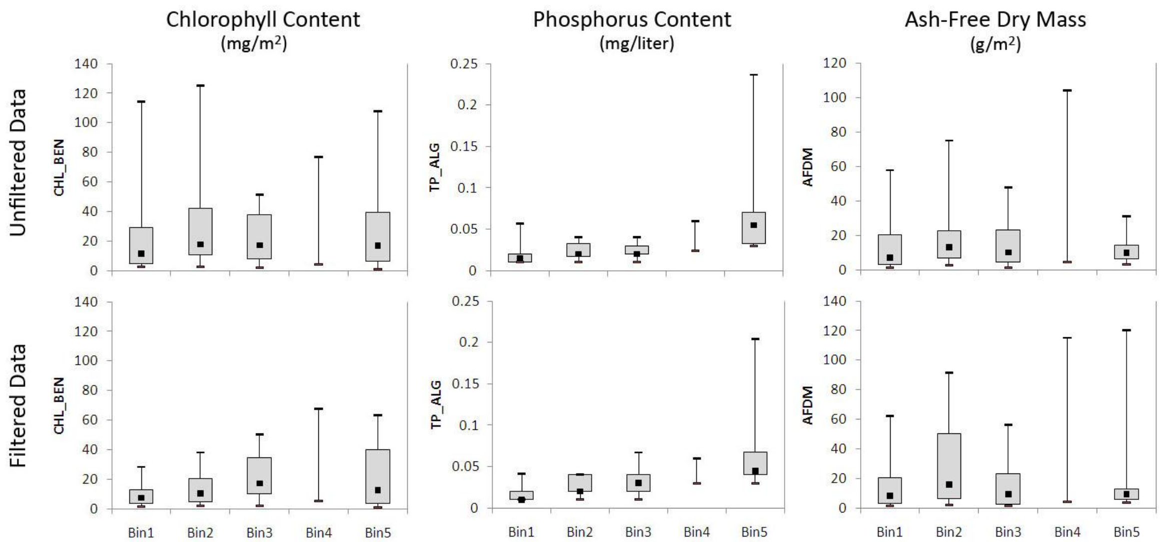


Figure 14. Periphyton chlorophyll *a* content (CHL_BEN), phosphorus content (TP_ALG), and ash-free dry mass (AFDM) distributions in the five nutrient bins. Unfiltered data includes all Virginia records for Strahler order 1-4 in Piedmont, Ridges, and Valleys. Filtered data include only those records with bank-related habitat parameter scores >10 and no karst geology at the sampling site. Conditions with the lowest observed TN and TP concentrations are represented by Bin1; those with the highest TN and TP concentrations are represented by Bin5. Bin3 has high TN and low TP; Bin4 has low TP and high TN. Solid square, median; box, IQR; whiskers, 5%ile – 95%ile. Bins with fewer than 10 samples are represented by a bar only. See **Table 11** for bin thresholds and **Table 12** for the median values of several water quality and habitat parameters.

Table 12. The median value in each bin of CHL_BEN, TP_ALG, AFDM, water column total phosphorus (TP_1y), total nitrogen (TN_1y), N:P ratio (TN:TP_1y), and specific conductivity (COND_1y), and Virginia's total habitat score (TotHabSc). See **Figure 14** heading for details. * less than 10 samples/bin.

Description	Bin1	Bin2	Bin3	Bin4*	Bin5	
TN	Very Low	Low	High	Low	High	
TP	Very Low	Low	Low	High	High	Units
Unfiltered (n=136)	*all records for Strahler stream order 1-4 in Piedmont, Ridges, and Valleys*					
CHL_BEN	7.5	20.6	19.8	11.3	21.6	mg/m^2
TP_ALG	0.015	0.02	0.02	0.04	0.055	mg/liter
AFDM	7.03	13.17	10.1	12.35	9.87	g/m^2
TP_1y	0.01	0.02	0.02	0.04	0.05	mg/liter
TN_1y	0.16	0.35	0.58	0.32	0.93	mg/liter
TN:TP_1y	11.8	18.0	43.0	6.3	16.9	-
TotHabSc	143	139	149	134	152	-
COND_1y	51	84	160	58	157	µmhos/cm
pH_1y	6.9	7.2	7.7	7.1	7.2	PSU
Filtered (n=74)	*only records with bank-related habitat scores >10 and no karst geology*					
CHL_BEN	7.5	10.4	17.1	11.3	12.5	mg/m^2
TP_ALG	0.01	0.02	0.03	0.04	0.045	mg/liter
AFDM	7.03	13.52	8.13	12.07	7.96	g/m^2
TP_1y	0.01	0.02	0.02	0.045	0.05	mg/liter
TN_1y	0.14	0.30	0.54	0.32	0.65	mg/liter
TN:TP_1y	11.8	14.0	27.2	6.3	11.8	-
TotHabSc	145	133	139	128	152	-
COND_1y	43	61	89	71	130	µmhos/cm
pH_1y	6.8	6.9	7.3	7.1	6.9	PSU

sample sizes when the data are filtered (**Table 11**) indicate that most of the confounded records came from bins 2 (Low TP_1y, Low TN_1y) and 3 (Low TP_1y, High TN_1y) which are the two most commonly occurring bins. AFDM did not show a distinct or consistent response in the exploratory RPART models to either nutrient concentration or to the possible confounding factors. The nutrient bins in **Table 11** were applied to the AFDM data even though the parameter did not exhibited nutrient breakpoints in the RPART trees. AFDM showed no consistent response to the nutrient bins (**Figure 14**).

5.3 Findings

The periphyton results should be considered preliminary because sample sizes in the nutrient bins are relatively small. The results apply only to streams (Strahler order 1-4) and to the combined Piedmont, Ridges, and Valleys bioregions. Other stream/river sizes and bioregions are under-represented in the data.

Chlorophyll *a* content and phosphorus content of periphyton appear to respond to ambient water column concentrations of TP_1y and TN_1y. Both metrics increase as TP_1y and TN_1y increase. Nutrient-related increases in chlorophyll *a* content are confounded or masked if habitat quality is poor or marginal and/or the sampling site is located on karst geology. Nutrient-related increases in phosphorus content occur regardless of the sampling site's habitat quality or geology.

Ash-free dry mass, or AFDM, does not seem to respond to increases in water column TP_1y and TN_1y. The result is not surprising in light of the fact that a relatively large component of the periphyton can be heterotrophic in well-shaded streams and small rivers. Periphyton heterotrophs, which include bacteria, fungi, protozoans, and small metazoans, extract much of their nutrients from submerged particles of organic matter such as leaf litter.

Periphyton chlorophyll *a* content above 100 mg/m^2 is suggested as a "nuisance" level in the 1999 USEPA Rapid Bioassessments Protocols (RBPs) for wadeable streams and rivers; the threshold reflects a general consensus in the literature (Stevenson and Bahls 1999). Eight of the 138 records met the "nuisance" threshold. All 8 records were associated with low TP_1y levels. Four of the eight classified in nutrient bin 3 (low TP_1y, high TN_1y) and the remainder classified in bins 1 and 2. By definition, bins 1 and 2 have low annual average concentrations of TN_1y (<0.43 mg/liter) and TP_1y (<0.04 mg/liter), and bin 3 has low TP_1y. All eight records were also associated with marginal/poor stream bank conditions and/or karst geology—conditions identified in the RPART summary as "primary" explanatory factors for periphyton chlorophyll *a* content. The habitat variables identified most often in the RPART model runs were bank stability (BANKS), stream bank vegetation (BANKVEG), width of the natural riparian vegetation (RIPVEG), and channel alteration (ALTER). The habitat findings suggest high flow scouring coupled with less stream shading due to loss of the bank and riparian vegetation are somehow related to increases in algal periphyton.

Conductivity was frequently identified as a "next best" factor explaining variability in the algal periphyton RPART model results. Conductivity is an indicator of the overall ion content of water and is generally higher in karst areas. It can be used to approximate water hardness which is the concentration of divalent cations. Calcium and magnesium, the dominant cations of hardness, precipitate soluble ortho-phosphate from the water column and lower the proportion of total phosphorus available to autotrophs. Summers (2008) noted that West Virginia rivers with water hardness greater than about 150 mg/liter had few or no instances of nuisance algal growth even though "ample phosphorus, nitrogen, and alkalinity were present." When hardness was below 100 mg/liter, nuisance growth could occur even when TP concentrations were fairly low. The low nutrient bins 1 and 2 have comparatively low COND_1y levels in both the filtered and unfiltered data (**Table 12**). These bins do not experience elevated and nuisance algal blooms when habitat conditions are suboptimal or optimal (filtered data); they can experience blooms if habitat conditions are poor or marginal (unfiltered data). Nutrient bin 3, characterized by high TN_1y and low TP_1y, has higher COND_1y levels with a median of 160 µmhos/cm in the unfiltered data and 89 µmhos/cm in the filtered data. The bin did not exhibit elevated nuisance or elevated algal levels, suggesting the concentration of hardness cations may be keeping water column ortho-phosphate concentrations low and limiting algal blooms. Bin 5 with the highest TN_1y and TP_1y concentrations experienced elevated and nuisance algal levels if habitat conditions were poor or marginal, even though its median COND_1y is 157 µmhos/cm; it did not when habitat conditions were suboptimal or optimal. Bin 5's TP_1y concentration is 2.5-fold higher than bin 3's, which might suggest that total phosphorus concentrations are saturating the capacity of the hardness divalent cations to precipitate ortho-phosphate. Overall, it appears that some constituents of conductivity—and most likely calcium and magnesium—confound autotrophic periphyton nutrient responses through their capacity to precipitate the ortho-phosphate component of total phosphorus. Whether the precipitated, settled phosphorus is truly unavailable to the periphyton is another question.

Stevenson and Bahls (1999) also suggest ash-free dry mass above 50 g/m^2 are nuisance levels. Nine of the AFDM records in the analysis data set met this threshold. All were from the filtered data, so bank-related habitat conditions were suboptimal or optimal and no site was located on karst geology. Furthermore, the records were associated with both high and low nutrient conditions: two in bin 1, three in bin 2 (the most frequently found bin), two in bin 3, one in bin 4, and one in bin 5. The findings substantiate the RPART results which suggest AFDM is not sensitive to water column TP_1y and TN_1y concentrations, at least in the Piedmont, Ridges, and Valleys, and it is not sensitive to bank-related habitat conditions or karst geology.

Despite the fact that periphyton can respond quickly to changes in their environment, this study's analysis approach of relating a periphyton sample to the averaged water quality from the prior year did produce a demonstrable nutrient response in two of the three periphyton metrics. In fact the strong relationship between TP_ALG and TP_1y suggests the latter metric might be a good indicator of the periphyton phosphorus content. Similar data analyses could be performed with other data sets to further explore and confirm the potential usefulness of the approach in evaluating water quality conditions for periphyton, either in the absence of a periphyton sample or when periphyton and water quality samples are separated in time and/or space. Hardness and alkalinity would be better environmental metrics to include in future analyses. It was not possible to include them in this analysis because of insufficient data.

In conclusion, chlorophyll *a* content and phosphorus content were the "bulk" periphyton metrics most sensitive to water column nutrient conditions while ash-free dry mass showed little if any response. Algal periphyton nutrient responses were observed when physical disturbances—namely, changes to the stream channel and bank and riparian vegetation—were minimal. Algal periphyton nutrient responses also appear to be complicated by high conductivity levels, some of which is probably natural due to karst geology.

6. Macroinvertebrates

Benthic macroinvertebrates are a diverse group of organisms that express a variety of morphological, behavioral, and feeding adaptations to lotic environments. Primarily, they are consumers of detritus, associated microbiota, and primary producers and are food for fish and higher trophic levels. They fill a variety of ecological niches and so can be used to measure and predict habitat suitability for aquatic life in most streams and rivers. Additionally, macroinvertebrates have a broad range of tolerances and sensitivities to different stressors and anthropogenic pollutants (Barbour *et al.* 1999). For these reasons, the use of benthic macroinvertebrates as indicators of overall water quality and habitat suitability is an accepted and widespread method for assessing the nation's streams. A variety of macroinvertebrate metrics can be calculated and used to quantify responses to stress. They measure taxonomic richness and diversity, feeding guilds, habit (e.g., swimmer, clinger), and tolerance values. Typically, collection of physical stream habitat data and several water quality parameters accompany benthic macroinvertebrate collections. This ancillary data is used in this study to partition the effects of confounding factors from the macroinvertebrate nutrient responses.

6.1 Data sources and metric selection

To improve the strength of the analyses and better control for confounders, we decided to combine data from outside the state of Maryland with Maryland-specific data. Sampling events outside of Maryland were selected if they were located in the bioregions found in Maryland, namely the Mid-Atlantic Coastal Plain, Southeastern Plain, Piedmont, Valleys, and Ridges bioregions. All data were extracted from the Chesapeake non-tidal stream benthic ("Chessie") database housed at the Chesapeake Bay Program (May 9, 2011 version). The database contains water quality measurements, physical habitat scores and macroinvertebrate taxa counts provided by 23 jurisdictional agencies in the Chesapeake Bay basin. Family-level macroinvertebrate metrics were calculated with the computer programs and associated lookup tables used in development of the "Chessie BIBI," or stream benthic index of biotic integrity for Chesapeake Bay basin (available from J. Johnson, CBP Living Resources Data Manager/Analyst). The database contains sampling events collected for various purposes and includes data flagged for quality assurance violations. Water quality data were filtered for quality assurance flags and for values reported as detection limits occurring above known ecological thresholds. For example, a series of TP detection limits reported as 0.129 mg/liter, a detection limit far above nutrient breakpoints commonly reported in the literature, was removed from the data set.

Numerical summaries were performed for each major classification scheme to determine if adequate numbers of key water quality and habitat variables were available for analysis. Samples sizes for each parameter were totaled by bioregion, and by bioregion and Strahler class. Sample sizes were also evaluated for events where basic water quality, nutrient chemistry, and physical habitat parameters were all collected with a benthic macroinvertebrate sample simultaneously. Incomplete data sets could be used to determine overall trends and provide support for relationships. To be used in the stressor-response analysis and to accurately evaluate the effects of confounders, sampling events were required to have the complete suite of parameters.

Macroinvertebrate data were provided by the jurisdictional agencies in the form of raw taxa counts. As would be expected, differences in macroinvertebrate sampling, processing, and taxonomical level of identification presented issues requiring adjustment. Differences in sampling protocol and equipment could not be addressed directly in the analysis but their effects were minimized. Taxonomical identifications were adjusted to the family level and sample sizes larger than 100 were sub-sampled to a standard 100-count using a Fortran rarefaction script (see description in Buchanan *et al*. 2011). A summary of the various methodologies can be found in USEPA 2011.

Fifty family-level metrics and the regional "Chessie IBIs" were calculated. They include a variety of tolerance, functional feeding guild, habit, diversity, and taxa group metric types. A description of the 50 metrics can be found in **Appendix B Table 1**.

A total of 8,789 sampling events with macroinvertebrate counts and water quality and physical habitat parameters comprised the dataset. The majority of events were collected in the Potomac River Basin (4,575), followed by Upper Chesapeake Bay drainages (1,341), the James River Basin (1,146), the Susquehanna River Basin (907), and the Lower Chesapeake drainages (818). To be used in the stressor-response analyses, all the extracted sampling events were required to have select stream habitat scores, pH, conductivity, DO, TN, and TP collected simultaneously with a benthic macroinvertebrate sample. A total of 1,712 sampling events met these requirements. As the project was not concerned with establishing current status, events dating back to 1993 were included.

6.2 Analysis Methods

Data analysis was organized into six steps of increasing specificity intended to eliminate the effects of confounding variables and identify nutrient thresholds for stream macroinvertebrate communities. The six steps are: 1) identify a subset of locations that are minimally disturbed by anthropogenic impact, 2) use data from the minimally disturbed locations (MDLs) to test and apply classification schemes, 3) explore nutrient-aquatic life relationships and identify stressor-response variables, 4) use recursive partitioning to filter out confounding habitat and water quality effects and identify nutrient breakpoints, 5) create categories (bins) with distinct nutrient ranges bounded by the breakpoints, and 6) link the nutrient bins to degradation of aquatic life.

6.2.1 Minimally disturbed locations (MDL)

A subset of samples from locations with relatively good water chemistry and habitat conditions was selected *a priori* and used to explore possible inherent differences in macroinvertebrate communities due to bioregion, season, and Strahler order. Samples from minimally disturbed locations (MDL) reduce variability in the biological response variables caused by anthropogenic impacts that would confound the influences of physiographic region, season and Strahler order. A sampling event had to meet all of the chemical and physical habitat selection criteria to be classified as minimally disturbed (**Table 13**). Overall, 947 sample events qualified for use as MDLs in the dataset. MDL samples were adequately represented across classification types with the exception of the Coastal Plain which had only 24 events that met the habitat criteria. Streams of 6th order and larger were underrepresented overall in the dataset (**Figure 15**).

Table 13. Criteria to identify minimally disturbed locations (MDLs) across all regions.

Parameter	Criteria
Specific Conductivity	$\leq$ 300 µS
pH	Between 6 and 9
Dissolved oxygen	> 5.0 mg/liter
Embeddedness	$\geq$ 16
Epifaunal Substrate	$\geq$ 16
Riffle Frequency/Quality	$\geq$ 16

The initial MDL data set included only observations with conductivity <300 uS, pH >6 and <9, DO >5.0 mg/liter, and scores for three habitat metrics—epifaunal substrate, embeddedness, and riffle frequency—equal to or greater than 16 of 20. Habitat and conductivity were included in the RPART models (below) and thresholds for filtering the data were later refined as lower breakpoints were discovered in the

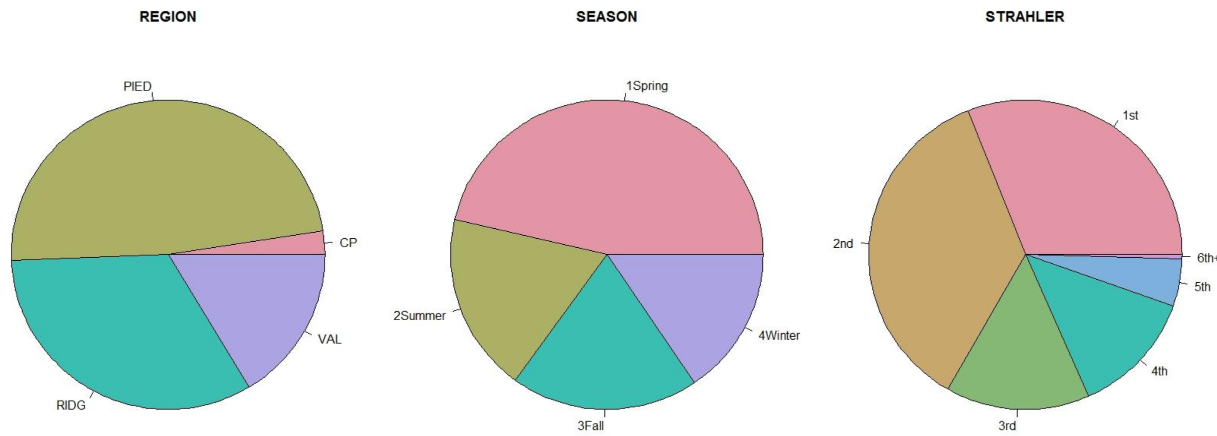


Figure 15. Distributions of minimally disturbed locations (MDLs) across season, region, and Strahler stream order.

recursive partitioning trees. The pH thresholds for filtering the data remained at 6 and 9 and the DO threshold remained at 5.0 mg/liter. Biological samples associated with pH <6 or >9 or dissolved oxygen < 5.0 mg/liter were removed from further analysis.

6.2.2 Classification

Macroinvertebrate assemblages are known to vary according to underlying abiotic characteristics such as soils, geology, elevation, stream size, and seasonality (Canton and Chadwick 1983, Vannote *et al.* 1980, Ormerod and Edwards 2006, Hawkins and Sedell 1981). Proper classification of data should minimize to a large extent the influence of these environmental factors on the macroinvertebrate variables, resulting in a clearer understanding of how macroinvertebrates response to nutrient stress. Several classification schemes were tested on the MDL data subset to determine which one could account for the most response variance. They were the Chesapeake Bay basin bioregions, Strahler stream order, and season.

The Chesapeake Bay bioregion classification scheme is described in detail in Buchanan *et al.* (2011). The bioregions relevant to this study are the Piedmont, Ridges, Valleys, Mid-Atlantic Coastal Plain and Southeastern Plain (**Figure 2**). The latter two bioregions were grouped and collectively call "Coastal Plain" for this analysis to increase sample sizes. A Strahler stream order assignment was obtained for each sampling location from the National Hydrography Dataset (1:100,000) stream layer. Seasons were assigned to each sampling event based on approximate distance from the equinoxes. Data collected in January, February, and March were assigned to Winter; April, May, and June to Spring; July, August, and September to Summer; and October, November and December to Fall. Classification testing began with visual investigation techniques such as box plots and continued with partitioning of variance using recursive partitioning and classification trees. The objectives of classification testing were to identify the most appropriate characteristics for data stratification, identification of appropriate classification assignment groupings, and to determine which classification criteria should be maintained in analyses in order to partition natural variability.

6.2.3 Exploratory analyses to identify stressor-response variables

Exploratory analyses were performed in R with the R Commander interface and the RExcel platform (Baier and Neuwirth 2007). Using the MDL samples, each of the 50 macroinvertebrate metrics were plotted as Tukey-style box and whisker plots, scatter plots, and/or histograms according to season, Strahler stream order, and bioregion and visually inspected for differences. Various measures of habitat and abiotic variables were also plotted by classification scheme to determine if differences existed in these factors and to inform potential breakpoints in the classification schemes. Quantile-quantile comparison plots were used to test the nutrient data for normality. TN and TP data were log-transformed for use in parametric procedures by adding 1.0 to the value and taking the base 10 log. Spearman-rank correlations and log-linear regressions were used to test for significant relationships between macroinvertebrates and nutrient parameters. Exploratory analyses continued using recursive partitioning analysis (see below).

A simple habitat quality index was constructed to quantify and further filter, if needed, the effects of in-stream and benthic habitat condition on the macroinvertebrate communities. Each jurisdictional agency employs a variation of the USEPA Rapid Bioassessment Protocol (RBP) to qualitatively and quantitatively measure aspects of the benthic, in-stream, and riparian habitat. While protocols differ slightly between jurisdictional agencies, many collect identical or similar habitat metrics. Three measures were selected from the available habitat metrics: epifaunal substrate quality, embeddedness, and either a riffle-frequency or riffle-quality metric. These three metrics are well represented across monitoring programs in the assembled dataset and they are assumed to have strong associations with macroinvertebrate fauna. Details explaining the comparability and translation of the habitat metrics can be found in one of several reports on the development and further refinement of the Chesapeake Non-Tidal Benthic Dataset (Buchanan *et al.* 2011). In order to be used as a numeric variable in the analyses, the three chosen metrics were combined into a simple habitat quality index (HQI) by summing the individual habitat scores for each sampling event. Combined, the index could have a possible range of 0 – 60.

6.2.4 Recursive partitioning

Recursive partitioning (RPART) analyses were performed in R with the R Commander interface and the RExcel platform (Baier and Neuwirth 2007). Data sets were loaded into the R dataframe from Excel and the packages RPART and PARTYKIT used to perform the analysis and diagram the recursive partitioning results. The PARTYKIT package provides a higher quality plotting of the constructed tree, complete with boxplot distributions of the response variable in each of the defined final nodes. For each metric, the RPART model makes binary splits of the data to create distinct nodes with less variance in both the macroinvertebrate metric and independent variables than the parent node. Splitting continues until either no further splits would result in decreased variance or until a predetermined minimum node size (usually n = 40) was reached. Care was taken to ensure proper filtering of the dataset for blanks so that independent stressor variables were evenly weighted.

In the exploratory analysis phase to assess the usefulness of different classification schemes, RPART models for 24 principal macroinvertebrate metrics were developed and applied to the initial MDL data set. Bioregion, season, and stream class (Strahler order 1^{st} – 3^{rd}, 4^{th} – 5^{th}, and 6^{th} +) were the independent variables. The model results show the Piedmont and Valley region always splitting together for all 24 tested metrics. In other words, variation in the metric scores between these two regions was insufficient to warrant division. Both regions have karst geology to varying degrees and similar topographies. These two bioregions could be grouped after further filtering for confounders.

To begin testing for possible nutrient breakpoints in the macroinvertebrate response variables, and to test for the influence of additional confounders, another series of RPART models was applied to the 24 macroinvertebrate metrics by bioregions. Each macroinvertebrate response variable in each bioregion was set against the following independent variables: specific conductivity (COND), stream class (STR_CLS), the habitat quality index (HQI), TN, and TP. The RPART model is:

[Macroinvertebrate metric] ~ COND + STR_CLS + SEASON + HQI + TN + TP

The initial round of recursive partitioning produced a suite of breakpoints for the model's variables which helped identify parameter levels likely to affect macroinvertebrates. Bioregion-specific conductivity thresholds for filtering the data were reset and applied; thresholds for filtering the habitat quality index were set and applied. The 24 macroinvertebrate metrics from each bioregion-specific, filtered data set were then run through the following RPART model:

[Macroinvertebrate metric] ~ SEASON + STR_CLS + TN + TP + $HQI_{filtered}$

where $HQI_{filtered}$ indicates the value is above-threshold. Season, STR_CLS and $HQI_{filtered}$ were left in the RPART model to check if any further thresholds appeared in these non-nutrient parameters. The summary function of the RPART application was called for each bioregion's 24 model runs. All the splitting parameters and their values, or breakpoints, in the summary outputs were exported to Excel and analyzed for most frequently reported nutrient breakpoints.

6.2.5 Nutrient bins

The binning approach was used to create a range of distinct nutrient categories, or bins, superimposed on environmental conditions that do not interfere with the macroinvertebrate community's responses to nutrients. The final runs of recursive partitioning models in the exploratory analysis above yielded a suite of possible breakpoints for both TN and TP. These nutrient breakpoints were used to construct the bins and test for biologically relevant thresholds of nutrient impairment. Histograms and pivot-tables were used to identify breakpoints reported most often in the RPART summaries. Frequently reported breakpoints for the low, middle, and high ends of the observed TN and TP ranges were used establish the boundaries of the nutrient bins. Data records with both TN and TP concentrations below the low boundary threshold classified as Bin A. On the other end, data records with TN and TP concentrations both above the high boundary threshold classified as Bin F. Bins B through E consisted of different combinations of low, medium and high TN and TP. Binning differed for each bioregion and was affected by the distribution and amount of data available in each bioregion. A minimum sample size of 10 determined whether there was sufficient data to construct a bin. Low abundance of TN data in the Valleys region prohibited construction of bins with adequate n sizes. The Valleys and Piedmont bioregions were combined based upon the RPART models that tested classification schemes (above) which did not split these two bioregions in any of the tested metrics. The data used in the bins was always filtered to minimize the confounding influences of conductivity, habitat quality, pH, and DO.

6.2.6 Quantifying biological condition

Degradation of habitat, food sources, and aquatic communities by nutrients in non-tidal streams and rivers occurs along a continuum. In order to quantify degradation of aquatic life, one must link certain nutrient levels with biological measures and values defined as being impaired. When using macroinvertebrates as indicators of impairment, this is typically done by scoring biological metric values against distributions observed in a defined reference dataset and combining the most responsive metrics into an Index of Biotic Integrity (IBI). The Maryland Biological Stream Survey (MBSS) of the state's Department of Natural Resources' program employs macroinvertebrate and fish IBIs to determine impairment in Maryland's non-tidal flowing waters. Unfortunately, their metric scoring thresholds for macroinvertebrates cannot be directly applied to the results of this study. Taxonomic resolution of the metrics differs between the two approaches with MBSS employing genus-level metrics and this work relying on family-level metrics. Additionally, the identification of reference (MDL) communities is different. This study developed an MDL condition based upon observed stressor breakpoints which do not always match those used by MBSS to identify its "reference" sites. Lastly, the spatial classifications of assessment differ. Maryland's Highlands region encompasses the Ridges and Valleys bioregions used in this study.

The Maryland MBSS approach for determining water quality with biological metrics, however, can be mimicked in this study. For each of its physiographic regions, Maryland identifies reference sites based on habitat, water quality, and land cover characteristics considered to be of high quality. Macroinvertebrate communities at these reference sites are assumed to have metric values that typify high quality. Percentiles of each metric's distribution of values at these sites are calculated and used to score the metric's values at all sampling locations as "most like reference" (score = 5), "somewhat like reference" (score = 3), and "least like reference" (score = 1). Macroinvertebrate metric scoring thresholds for this study were derived in a manner similar to Maryland's but using a refined set of MDLs instead of MBSS reference sites. **Table 14** lists the criteria for identifying the refined MDL sampling events.

The 10th, 50th, and 90th percentiles were calculated for all 50 macroinvertebrate metrics from the population of refined MDL sampling events. For metrics that respond positively to anthropogenic stress, such as the percent tolerant individuals in a sample, values occurring at or above the 90th percentile of the refined MDL data were scored "1," values falling between the 50th and 90th percentiles were scored "3," and values falling below the 50th percentile value were score "5." For metrics that respond negatively to anthropogenic stress, such as the percent sensitive individuals in a sample, values below the 10th percentile of the refined MDL data were scored "1," values falling between the 50th and 10th percentile values were scored "3," and values falling above the 50th percentile value would score as "5". This 1-3-5 scoring approach mirrors the method used by Maryland MBSS for scoring macroinvertebrate metric values and many of the calculated percentile values are very similar, despite being drawn from populations identified with different approaches (See 6.4.2 for a comparison of scoring thresholds for metrics used in the MBSS regional BIBIs). The percent of samples in each nutrient bin that scored "1" is considered to be the measure of degradation in each bin. Values of the 10th, 50th, and 90th percentiles used as scoring thresholds for the 50 metrics used in this study are listed in **Appendix B Table 2**.

Table 14. Criteria used to refine the MDL sampling events for the purpose of establishing metric scoring thresholds.

Region	pH	DO	COND	HQI
Coastal Plain	6 - 9	> 5 mg/L	< 200	> 40
Piedmont	6 - 9	> 5 mg/L	< 200	> 50
Valley	6 - 9	> 5 mg/L	< 342	> 50
Ridge	6 - 9	> 5 mg/L	< 242	> 50

6.3 Results

6.3.1 Classification

Visualization

With a few exceptions, the family-level macroinvertebrate metrics used in this study display low variance across season and stream size classes in MDL conditions (see **Appendix B Table 3** and **Table 4**). For example, the classic "taxa richness" metric did not vary across Strahler stream order or season (**Figure 16**). Those metrics that did vary by Strahler order and season were measuring the relative

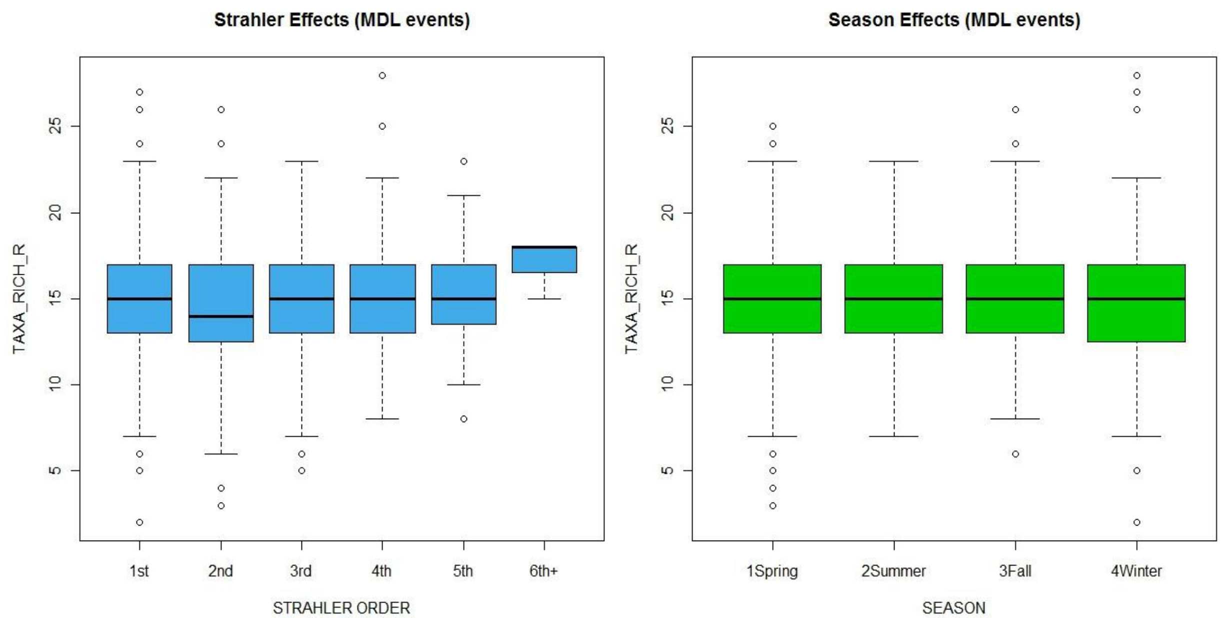


Figure 16. Taxa richness (100-count sample) in MDL conditions, plotted by Strahler stream order and season.

abundance of functional feeding groups, habit types, and certain specialist taxa. For example, the percentages of filterers, scrapers, and gatherers display seasonal differences and the percent of shredders, and more specifically stoneflies, display an association with headwater streams. These shifts are expected in the context of the River Continuum Concept. Gatherer abundance will decrease and scraper and filterer abundances increase in response to the greater availability of periphyton and phytoplankton food sources during the summer and fall seasons (**Figure 17**).

Similarly, Plecoptera (Stoneflies) are predominantly shredders and rely on coarse organic material, usually in the form of leaves that are contributed to upland, forested headwater streams. For this reason,

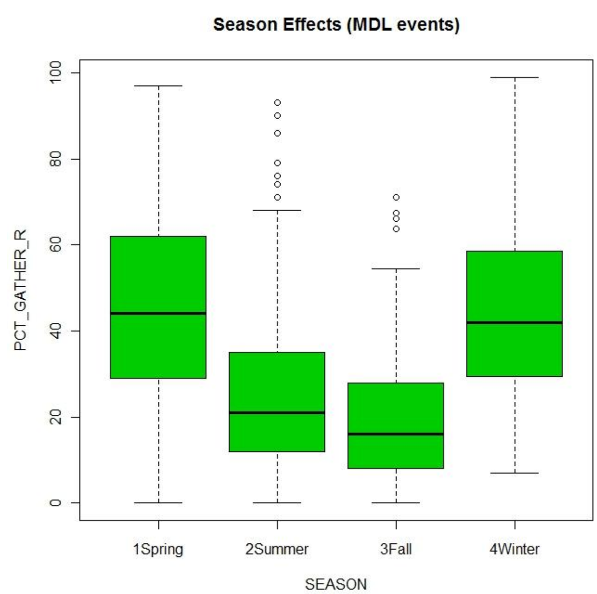


Figure 17. Percent gatherers (100-count sample) in MDL conditions, by season.

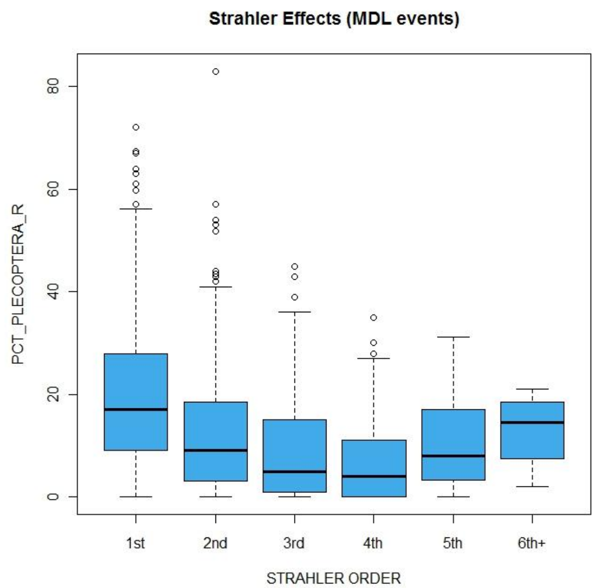


Figure 18. Percent Plecoptera (100-count sample) in MDL conditions, by Strahler stream order.

abundances are expected to be higher in low order, headwater streams (**Figure 18**).

Visualizing the macroinvertebrate metrics by bioregion revealed marked differences among the macroinvertebrate assemblages. The combined Coastal Plain bioregions support higher proportions of tolerant and non-insect taxa compared to streams in the Piedmont, Ridges, and Valleys bioregions. Higher proportions of Diptera taxa and fewer EPT taxa together with more tolerance taxa contribute to significant differences among calculated diversity and tolerance indexes such as the Shannon-Weiner and Beck's Biotic Index (**Figure 19**). Coastal Plain streams are typically low gradient, with sandy substrate which is less than optimal habitat for benthic invertebrates which prefer more stable habitat such as cobble and boulder. Similarly, yet less pronouncedly, is the differentiation of the Ridges from the other bioregion classifications. Measures of relative macroinvertebrate health are more positive in the Ridges, revealing higher abundances of sensitive taxa such as EPT. The Ridges bioregion, being unsuitable for farming and development, is often less disturbed by anthropogenic sources and more likely to have intact riparian buffers. (Historically, these same areas were heavily impacted by logging and the subsequent topsoil erosion.) The upland streams of the Ridges are also of higher gradients and typically have adequate to optimal substrate for benthic habitat. The Piedmont and Valley bioregions did not separate in any of the exploratory RPART models suggesting that the regions could be combined for analyses.

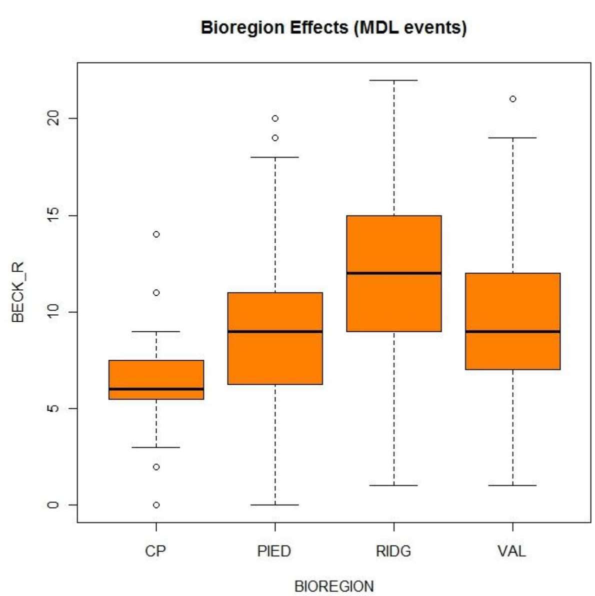


Figure 19. Beck's Biotic Index (calculated from 100-count samples) in MDL conditions, by bioregion.

Light availability is an important limiter of primary productivity; however, actual measures of this confounder were rare in the dataset. To test if Strahler stream order could be used to account for light availability, the estimated percent shading of streams with good riparian buffer scores ($\geq$16 of 20) was plotted by stream order. No data was available above fourth order, however it was clear that percent shading decreased with increasing stream order (**Figure 4**).

Recursive partitioning

In the example shown in **Figure 20**, variability in the percentage of tolerant invertebrates in the MDL data is partitioned to minimize variability in the biological metric as well as bioregion, season, and Strahler order. The first division splits the Coastal Plain from the other regions, indicating the biotic and abiotic features of its records differ most from other records. Each branch then splits again, with the Coastal Plain samples dividing into Strahler orders 1st-4th and 5th-6th. The Piedmont, Ridges, and Valleys branch splits first on season, with fall and summer having low percentages of tolerant. The spring and winter samples split a final time into a Ridge node and a Piedmont and Valley node. The relative distribution of the percent of tolerant organisms in each terminal node is then graphically displayed in a boxplot.

Results of the recursive partitioning for the 24 principal macroinvertebrate metrics in MDL conditions are summarized in **Table 15**. For each metric, an "X" denotes the primary or first split, "#" denotes a second-level split, and "*" a third-level split. The results mirror the observations of the box-plots, reinforcing the conclusion that the bioregion classifications explain a lot of the variance observed in the calculated benthic macroinvertebrate metrics. In all, 18 of the 24 tested metrics were best classified by the bioregion assignments. Five of the remaining six metrics were functional feeding group metrics for which season was shown to have the strongest explanatory effect. These results indicated that in order to best evaluate stressor-response relationships between nutrients and benthic macroinvertebrates, the data should be classified and analyzed separately by bioregions. This removes the confounding effects of regional differences and decreases overall variability in the response variables. Since season and stream size effects are sometimes observed, it was decided to retain these as variables in multivariate analyses.

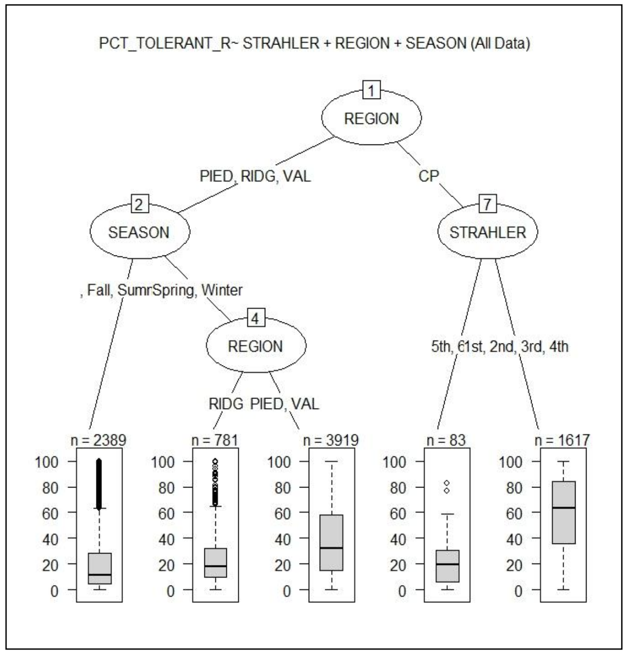


Figure 20. PARTYKIT plot of the classification tree built by the RPART function for the metric %Tolerant to inform classification of the stressor-response data.

Table 15. Results of the classification testing for macroinvertebrate metrics (MDL conditions).

Metric	Bioregion	Season	Strahler
ASPT_MOD	X, #		
BECKS	X, #		
EPHEMEROPTERA TAXA	X, #	#, *	
EPT TAXA	X, #		
FAMILY HILSENHOFF	X, #	*	
NON-INSECT TAXA	X, #		
%_CLINGER	X	#	
%_COLLECTOR	#	X	
%_DOM3	#, *	X	#
%_EPHEMEROPTERA	X, *		#
%_EPT	X, #	*	
%_FILTERER	#	X	#
%_NON-INSECT	X, #		
%_PLECOPTERA	X, #	*	#
%_SCRAPER	*	X	#
%_SENSITIVE	X, #	*	
%_SHREDDER	#		X
%_SWIMMER	X	#	*
%_TOLERANT	X, *	#	#
PLECOPTERA TAXA	X, #		#
SENSITIVE TAXA	X, #		
SHANNON-WEINER	#, *	X	#
TAXA RICHNESS	X		#
TOLERANT TAXA	X, #		

6.3.2 Exploratory analysis

Visualization

A variety of plotting techniques were used to visualize and investigate patterns in nutrient variables in the entire, unfiltered data set. Histograms were used to examine nutrient concentrations in a particular classification scheme and to detect outliers missed during earlier quality assurance checks. In all regions, the distributions of total nutrient concentrations were heavily skewed, with the majority of values being low and higher concentrations decreasing in frequency (**Appendix B Figure 1**). Across all regions, total phosphorous data was more heavily skewed than was total nitrogen. Q-Q plots were used to test the normality of the total nutrient distributions within each region. Overall, the quantile comparisons did not show good agreement with a theoretical normal distribution. In order to increase normality for parametric tests, data were log-transformed by adding 1.0 to the original measurement and taking the base-10 log of the value. Log-transformed TN values were found to be more normally distributed than the original measurements (**Figure 21**). However, log-transforming the TP data did not increase the normality in any region.

Tukey-style box plots were used to investigate differences in nutrient concentrations across bioregions and across Strahler stream order. Bioregions differed significantly in their stream nutrient concentrations, with the Piedmont and Valley streams having higher water column TN levels than either the Ridge or Coastal Plain streams (**Figure 22**), and Coastal Plain streams having higher water column TP levels than the other bioregions. Consistent patterns were not observed in median TN and TP concentrations across 1st – 4th order streams, although the upper percentile concentrations do tend to decline with increasing stream order.

Correlation and regression analyses

Spearman-rank correlations and simple linear regressions were used to test for significant stressor-response relationships between nutrients and macroinvertebrate metrics in the entire, unfiltered data set. Spearman-rank correlations were made using non-transformed total nitrogen and total phosphorous concentrations and the 50 candidate benthic metrics while simple linear regressions were

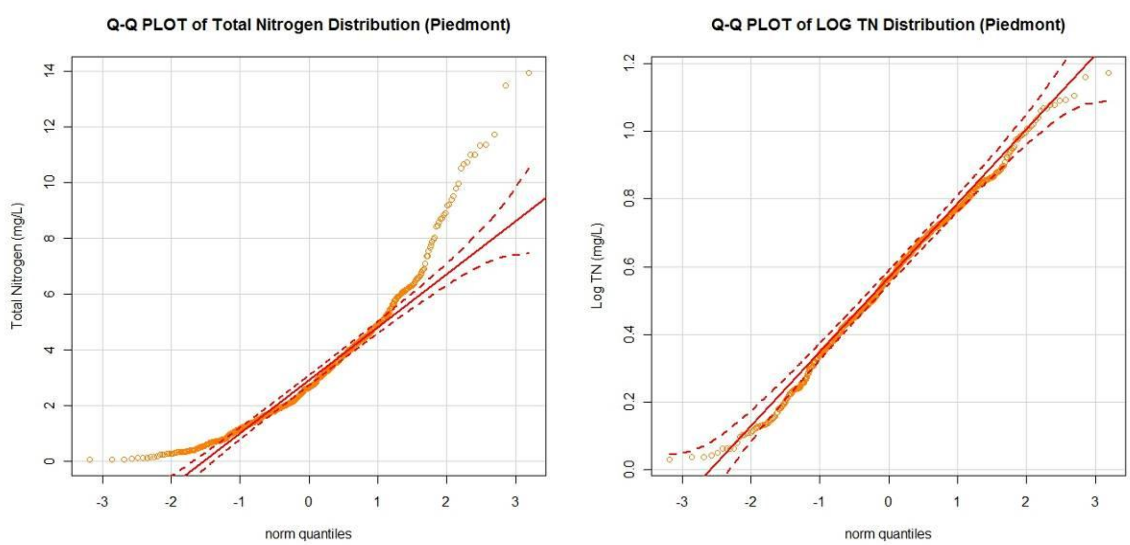


Figure 21. Quantile-quantile comparison plots of total nitrogen and log-transformed total nitrogen data in the Piedmont region. All available data are used.

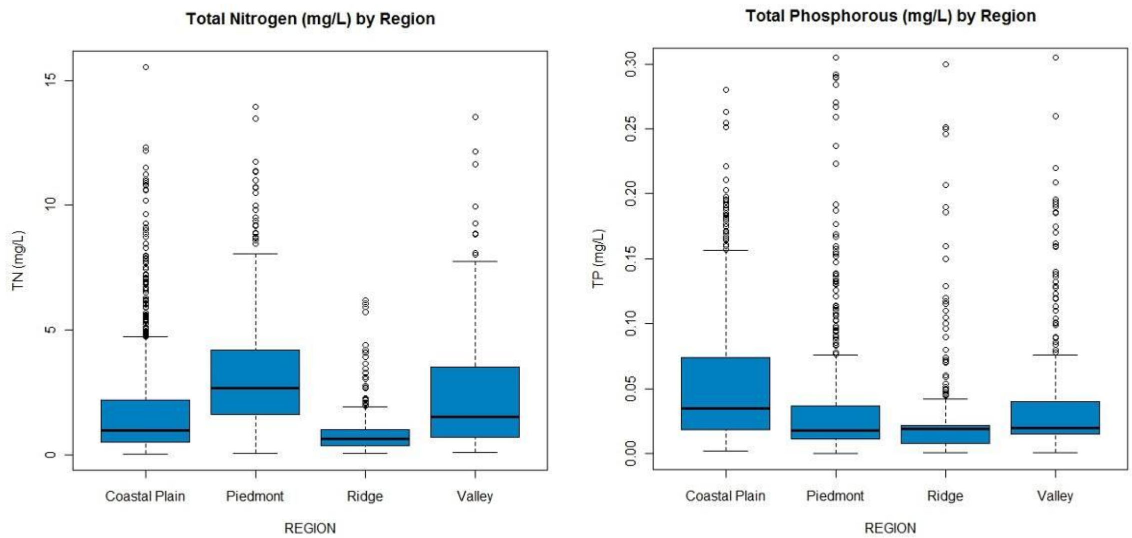


Figure 22. Total nitrogen and total phosphorous distributions across regions. TP was plotted without outliers above 0.30 mg/L in order to better visualize the distribution. All available data are used.

performed using the log-transformed nutrient concentrations. The resulting correlation coefficients, p-values measuring the significance of the linear regressions, and R^2 values were ranked to identify those metrics with the most significant associative response to nutrients (Spearman-rank results are given in **Appendix B Table 5**). Generally, macroinvertebrate metrics responded as would be predicted, with those metrics that respond negatively to increasing stressors having negative correlation coefficients, and *vice versa*. Not all benthic metrics had significant relationships to one or both nutrient measures, and low R^2 values overall indicated that nutrients weren't able to account for much of the variance in the macroinvertebrate metrics.

Scatter plots were used to plot the results of the simple linear regression analyses. Like the correlation and regression analyses, the scatter plots suggested that while there was an observable nutrient-aquatic life response, other variables have strong confounding effects. In **Figure 23**, Hilsenhoff's Family-level Biotic Index is plotted against TN for the Valley region and a significant but moderate relationship can be seen. Index values are increasing with nutrients, indicating degradation, but there is obvious scatter of the response values around the regression line. Another commonly observed pattern that emerged was the "classic wedge" such as the example shown in **Figure 1** for the Beck's Biotic Index *versus* TP in the Piedmont bioregion. A range of index values occur at low TP concentrations, indicating stressors other than nutrients are impacting the biota, while high

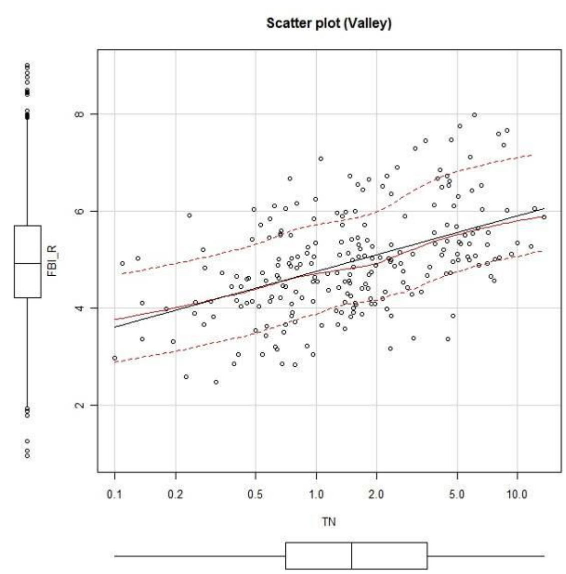


Figure 23. Hilsenhoff Family-level Biotic Index (FBI) plotted against TN in the Valleys bioregion.

Table 16. List of 24 candidate macroinvertebrate response metrics

ASPT MOD
BECK'S BIOTIC INDEX
EPHEMEROPTERA TAXA
EPT TAXA
HILSENHOFF FAMILY-LEVEL BIOTIC INDEX
NON-INSECT TAXA
% CLINGER
% COLLECTOR
% DOM3
% EPHEMEROPTERA
% EPT
% FILTERERS
% NON-INSECT
% PLECOPTERA
% SCRAPER
% SENSITIVE
% SHREDDER
% SWIMMER
% TOLERANT
PLECOPTERA TAXA
SENSITIVE TAXA
SHANNON-WEINER
TAXA RICHNESS
TOLERANT TAXA

index values do not occur at high TP concentrations, indicating either TP is a significant cause of the observed impairment or confounding variables are increasing in concert with TP.

Choosing candidate stressor-response variables

From the original suite of 50 macroinvertebrate metrics, 24 were chosen for use in multivariate analyses, based primarily upon the strength and the consistency of their observed nutrient-relationships in the exploratory analyses of the entire, unfiltered data and secondarily on our objective of maintaining a variety of metric types. Evidence from the Spearman-rank correlations, simple linear regressions, and visual plotting were all considered in choosing the final list of metrics. Not all tested metrics showed a consistent or predictable response to nutrients and were subsequently removed to lessen computational workload. The final set of 24 metrics chosen for multivariate analyses are listed in **Table 16** and consist of a variety of richness, tolerance, habit, and functional feeding group metrics.

The nutrient parameters of TN and TP were chosen as the primary stressor variables for several reasons: they were well reported in the collected datasets; they may give a more accurate picture of overall trophic state of the stream; most candidate numerical nutrient criteria are set for TN and TP; and lastly, other nutrient species thresholds could be easily investigated once a analysis framework had been established.

The habitat quality index (HQI) created from the three habitat metrics used to identify MDL conditions was employed as the RPART model habitat variable. The HQI appears to adequately represent the influence of habitat on the macroinvertebrate response metrics. Most macroinvertebrate metrics associated with above-threshold HQI scores are clearly separated from those associated with low, or degraded, HQI scores across all bioregions. Two example plots are shown in **Figure 24**. Slopes of the linear regression relating nutrients and macroinvertebrates in the scatter plot are significantly separated when the data are plotted by habitat assignments, indicating that habitat is acting as a severe confounder of the macroinvertebrate nutrient responses.

6.3.3 Confounding factors

Before macroinvertebrate responses to nutrients could be explored, sampling events associated with environmental conditions known to impair biological communities or confound nutrient responses were removed, or filtered, from the analysis data set. In the suite of information associated with each sampling event, habitat quality, conductivity, dissolved oxygen, pH, bioregion, season, and stream size were considered to be the abiotic factors most likely to confound biological nutrient responses. Other potentially confounding factors such as toxic chemicals, flow, food availability, and predation are not consistently monitored so their influence could not be examined.

Efforts to further filter the analysis data set began with the MDL thresholds for dissolved oxygen and pH (**Table 13**). These thresholds for filtering the data remained at DO <5.0 mg/liter and pH <6 and >9. Dissolved oxygen and pH levels are currently regulated by Maryland and exceedances are known to

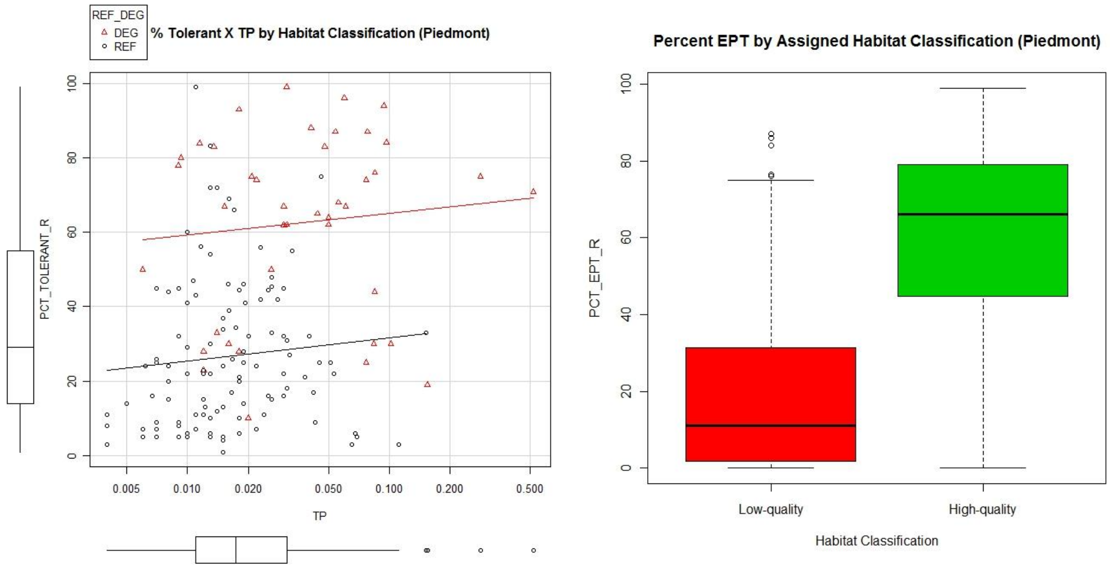


Figure 24. Selected plots showing the separation in response metric values between the high and low-quality habitat assignments. The example on the left is a scatter plot of the percentage of the sample comprised of tolerant invertebrates along a TP gradient in the Piedmont region. The box plot on the right shows the percentage of EPT taxa in a sample between high and low-scoring habitats in the Piedmont.

stress macroinvertebrate communities. Successive recursive partitioning models were then applied to the pH- and DO-filtered data to identify specific conductivity and HQI thresholds.

Stream order and season were retained as classes in the RPART models. Models were run on the 24 candidate biological metrics in each of the four bioregions for a total of 96 model runs. RPART model results indicated specific conductivity was a strong predictor of macroinvertebrate metric scores followed by habitat quality. Specific conductivity topped HQI as the primary explanatory variable in 65 of the 96 model runs. For example, a first split at the conductivity value of 330.6 µS best minimized the variance in both the Beck's Biotic Index score and the independent variables in the Valleys analysis data set. Additional splits using the habitat quality index and a lower conductivity threshold were also observed. In each split, either a higher HQI or a lower conductivity value led to overall higher Beck's Index scores.

Noticeable differences between bioregions were found in the conductivity thresholds identified by the RPART models (**Table 17**). Thresholds that split lower metric scores from higher scores were higher in the Valley than in the other three bioregions. This difference can be attributed to Karst formations found in the Valley region, which would contribute Ca+ and CO_3- ions from the prevalent limestone geology present there, giving the Valley a naturally higher baseline of conductivity that is not necessarily a stressor of benthic communities.

In order to account for conductivity as a confounder, the identified thresholds were averaged by region to find the value above which conductivity was presumed to be a strong confounding variable. These average values varied by region and are reported in **Table 17**.

The HQI breakpoints observed in the RPART trees differed by bioregion. Filtering values were based upon values observed in the recursive partitioning trees in order to remove samples with habitat values

Table 17. Observed breakpoints where specific conductivity was chosen as the primary explanatory variable in the RPART trees.

Metric	Ridges	Valleys	Piedmont	Coastal
ASPT MOD	217	330	235	
BECK'S BIOTIC INDEX	245	331	209	132
EPHEMEROPTERA TAXA	435	396	210	
EPT TAXA	245	331	235	
HILSENHOFF BIOTIC INDEX	248		206	132
NON-INSECT TAXA	30		132	
% CLINGER	248		191	
% COLLECTOR	71	313	209	
% DOM3	249	441	209	132
% EPHEMEROPTERA	271		133	148
% EPT	248		206	
% FILTERERS				
% NON-INSECT	219			
% PLECOPTERA		165	251	128
% SCRAPER	269			
% SENSITIVE	248	333	190	120
% SHREDDER			247	
% SWIMMER	428	287	88	128
% TOLERANT	249		206	
PLECOPTERA TAXA	182	165	222	132
SENSITIVE TAXA	245	331	209	132
SHANNON-WEINER	249	441	209	264
TAXA RICHNESS	249	411	276	294
TOLERANT TAXA			125	
AVERAGE	242.3	328.5	199.9	130.9

that resulted in very poor biological metric values. For example, low HQI values identified poorly scoring Beck's Index values.

6.3.4 Recursive partitioning to identify nutrient breakpoints

The final round of recursive partitioning was performed on confounder-filtered data sets using the following RPART model:

[Macroinvertebrate metric] ~ SEASON + STR_CLS + TN + TP + HQI

Bioregions were analyzed separately at first. Categories for season (Spring, Summer, Fall, Winter) and Strahler stream order (1^{st} – 3^{rd}, 4^{th} – 5^{th}, 6^{th}+) were retained in the model because of their known importance to certain metrics**.** Sampling events with HQI, pH, SPCOND and DO values outside the thresholds listed in **Table 17** were filtered from the analysis data sets in a manner similar to the refined MDL pool. However, HQI was retained in the model to confirm the absence of further habitat-related confounding influences. Each record had to have all parameters to be included in the analysis. The final round of recursive partitioning models produced a suite of observed nutrient breakpoints across bioregions and macroinvertebrate response metrics. Examples of RPART model output from three bioregions appear in **Figures 25 – 27.**

6.3.5 Nutrient bins

The most frequently observed low, middle and high breakpoints produced by the RPART analyses were used to construct nutrient bins and quantify the combined effect of nitrogen and phosphorus on the macroinvertebrate response variables. **Table 18** lists the parameters and values that were used to filter the analysis dataset for confounding variables and **Table 19** lists the selected breakpoints used to construct the nutrient bins. The resulting bins and bin thresholds differ by region because they were empirically derived from the available data. In the Ridges bioregion, for example, there were very few instances of comparatively high concentrations for both nitrogen and phosphorous in the filtered dataset. Hence, TN and TP thresholds in the bioregion's "Very High" bin (F) are lower than those in other regions. In the Coastal Plain, nutrient concentrations overall tend to be higher, so the "Very Low" bin (A) has TN and TP thresholds that are higher than in other tested bioregions. Additionally, filtering to remove sampling events with confounding stressor values removed many events where high nutrients were also recorded and contributed to low n-sizes in some of the higher bins. This was especially true in the Ridges where the control of specific conductivity was iteratively relaxed in order to meet n size requirements in the highest bin.

Plotting of the nutrient bins revealed patterns of association between the benthic response metrics and the binned nutrient concentrations. In the Coastal Plain and the combined Piedmont/Valleys regions, clear patterns of nutrient-responses were evident, with higher nutrient bin assignments resulting in overall poorer benthic metric scores. In the Ridges, where nutrient concentrations were lower, it is evident that the lower nutrient bins may be reflecting oligotrophic conditions, and patterns of degradation may be more closely linked to fluctuations in the N:P ratio and its effects on food quality than on eutrophic conditions.

If the MBSS scoring approach is applied to the filtered data set, values below the 10th percentile (for metrics that decrease with nutrient stress) or above the 90th percentile (for metrics that increase with nutrient stress) would be scored "1," indicating possible degradation. Values above or below the 50th percentile (for metrics that decrease or increase with nutrient stress, respectively) are scored as"5," indicating high quality. For many biological metrics in the Coastal Plain and Piedmont/Valleys, low scoring values are significantly more

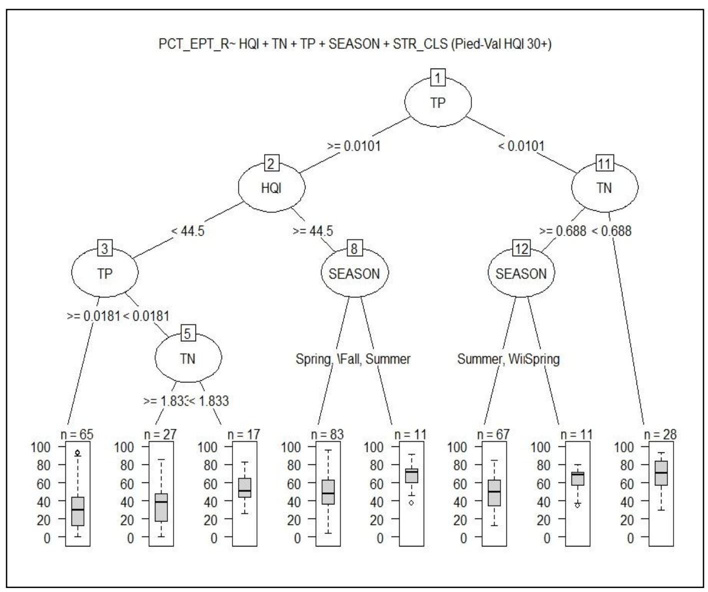


Figure 25. Piedmont+Valleys RPART model and tree for percent EPT. Total phosphorous was the primary splitting variable, dividing the samples around a value of 0.010 mg/L with overall higher percent EPT observed in lower TP concentrations. Data in the low TP concentrations split further on TN (0.688 mg/liter) and Season. Data in the high TP concentrations split further on HQI, then on Season, a higher TP breakpoint (0.0181 mg/liter), and a higher TN breakpoint (1.833 mg/liter).

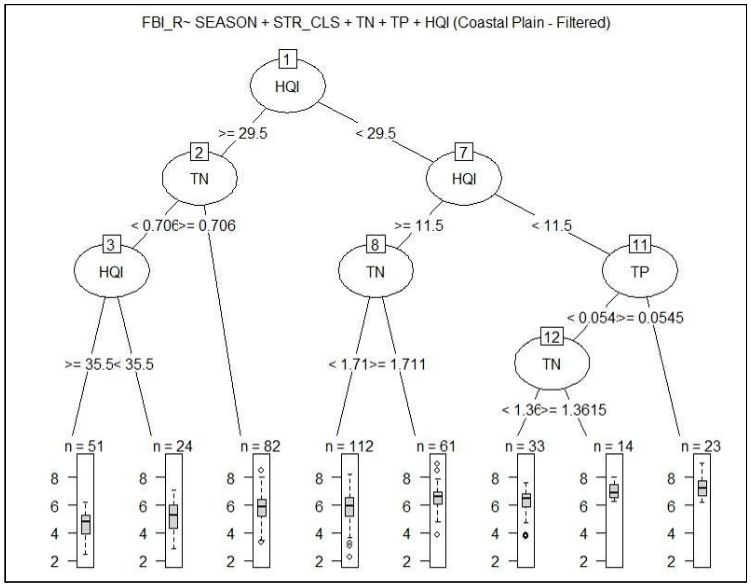


Figure 27. Coastal Plain RPART model and tree for the Hilsenhoff Family-Level Biotic Index (100-count samples). The FBI_R responds positively to stress, so a higher score is indicative of impairment. In this example, increasing nutrients and decreasing habitat quality lead to higher FBI scores.

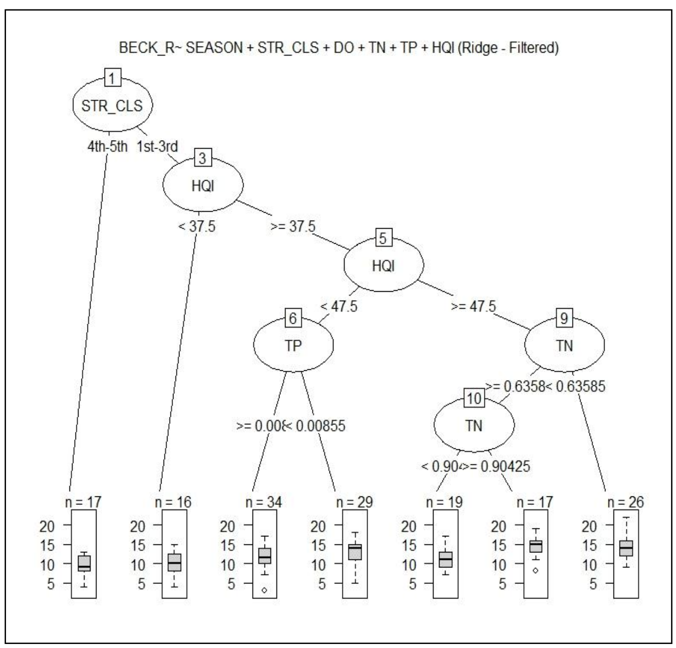


Figure 26. Ridges RPART model and tree for Becks Index. Seventeen 4^{th}-5^{th} order Ridge streams and small rivers were split at the first node from the other streams. HQI split the second and third nodes. Breakpoints for TP (0.086 mg/liter) and TN (0.64 mg/liter and 0.90 mg/liter) were observed in moderate to high habitat quality samples. Lower TP values translated to higher Beck's Index scores. A consistent TN response was not observed and although TN was selected by the model as the best splitter variable, it may be reflecting the effects of factors not accounted for in the model.

Table 18. Thresholds used to filter data set prior to nutrient breakpoint analyses. * Data was filtered for conductivities above 200 µS, despite the lower average breakpoint defined by RPART analysis for the Coastal Plain. ** Conductivity values for filtering were incrementally raised from the RPART identified threshold of 242 µS to 282 µS in order to obtain 10 or more samples for all nutrient bins to test for nutrient effects.

Variable	Ridges	Valleys	Piedmont	Coastal Plain
pH	5.5 – 9	6 – 9	6 – 9	6 – 9
DO	≥ 5.0	≥ 5.0	≥ 5.0	≥ 5.0
Conductivity	< 282**	< 329	< 200	< 200*
HQI	≥ 35	≥ 30	≥ 30	≥ 25

Table 19. Nutrient breakpoints used to establish nutrient bins in the regions. * Bin A is not included in Bin B; ** Bin F is not included in Bin E.

	Ridges			Coastal Plain			Piedmont/Valleys		
BINS	TN	TP	n	TN	TP	n	TN	TP	n
A: Very Low	< 0.22	< 0.006	9	< 0.82	< 0.018	23	< 0.64	< 0.013	27
B: Low*	< 0.50	< 0.01	34	< 1.52	< 0.061	86	< 1.65	< 0.029	72
C: High N / Low P	≥ 0.50	< 0.01	44	≥ 1.52	< 0.061	47	≥ 1.65	< 0.029	146
D: High P / Low N	< 0.50	≥ 0.01	10	< 1.52	≥ 0.061	11	< 1.65	≥ 0.029	13
E: High**	≥ 0.50	≥ 0.01	50	≥ 1.52	≥ 0.061	11	≥ 1.65	≥ 0.029	32
F: Very High	> 0.91	> 0.019	13	-	-	-	≥ 2.65	≥ 0.045	12

frequent in the High and Very High nutrient bins. In the Ridges, the lowest scores occur most often in the lowest nutrient bin (A), the high TP bin (D) and the highest nutrient bin (F). **Figures 28 - 30** present the binning results for the %Sensitive metric in the three regions. In each plot, the green line indicates the 50th percentile and the red line indicates the 10th percentile of the values observed in the refined MDL pool.

In the Coastal Plain, the percentage of macroinvertebrates in the sample considered sensitive (tolerance values ≤ 3) begins to decline in bins C and D, the High TN and High TP bins, respectively, and drops further in the highest nutrient bin (E) where both TN and TP are enriched (**Figure 28**). Following the MBSS scoring approach, the percentage of samples that would score as "1" is highest in bin E. A sample scores "1" if less than 2% of the sample is comprised of sensitive individuals. The number of samples that would score "5", with percentages of sensitive macroinvertebrates > 20.65% in each sample, is highest in bins A and B, indicating that concentrations of TN and TP below the breakpoints used to delineate the bins are protective of aquatic life. As concentrations of either TN or TP exceed these breakpoints, and especially when both are exceeded, the chance of scoring "1" increases. This pattern of nutrient-related degradation in benthic metric scores was generally consistent, and was observed in most of the 50 tested metrics and the "Chessie IBI". **Appendix B Figure 2** displays 12 sample plots for each of the three tested regions.

In the combined Piedmont/Valleys region, the observed pattern was very similar; however, the point at which significant change in responses of the metric values appears earlier (**Figure 29**). In this region, the most significant change occurred between bins A and B, with metric values indicating degradation beginning with TN and TP concentrations above the breakpoints used to define bin A. This plot also

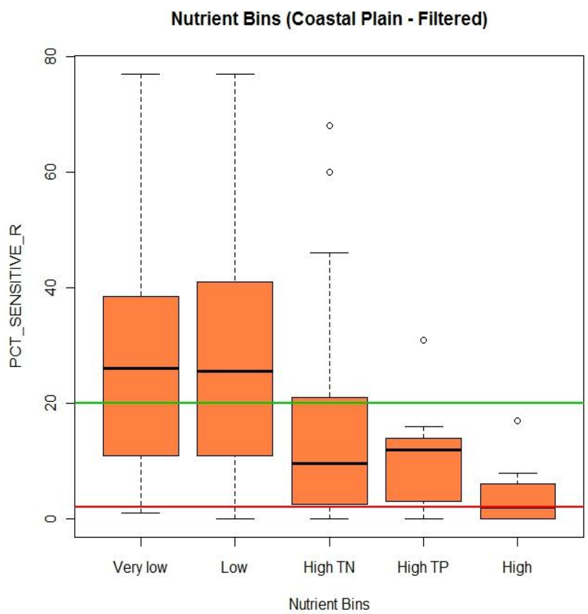


Figure 28. Percent Sensitive organisms plotted by nutrient bin in the Coastal Plain bioregion. The green horizontal line indicates the value of the 50^{th} percentile and red line is the value of the 10^{th} percentile observed in the refined MDL dataset.

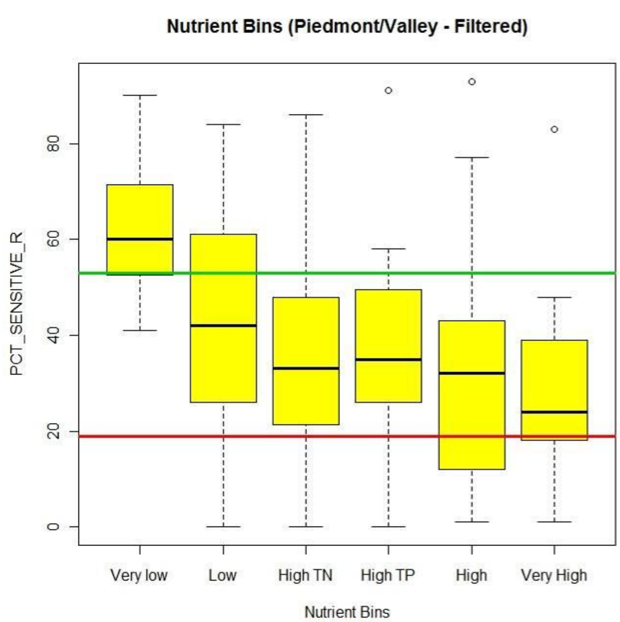


Figure 29. Percent Sensitive organisms plotted by nutrient bin in the combined Piedmont and Valley bioregions. See **Figure 28** heading for details.

helps to illustrate a further point: nutrient-related impacts occur along a continuum with no clearly defined threshold of impairment. Nutrient-related impacts are likely to increase until a system is saturated. Chances of exceeding aquatic life impairment standards can be expected to increase with increasing TN and TP concentrations, and the binning approach helps to quantify this expectation. In this example, Piedmont/Valleys streams, barring impacts from other stressors, would be expected to meet aquatic life standards 100% of the time if TN and TP concentrations were less than 0.64mg/liter and 0.013 mg/liter respectively, but would be expected to fail 36.4% and 30.8% of the time at concentrations above those used to construct bins E and F (See **Appendix B Table 6** for percent exceedances for the 50 tested metrics).

The Ridges region, as stated before, had overall lower TN and TP concentrations and the constructed bins were therefore similarly lower, as guided by the data distribution and observed RPART breakpoints. As could be expected, the observed metric responses did not follow a pattern consistent with the other tested regions. There was, however, an observable pattern that was repeated for multiple metrics (**Figure 30**). Contrary to the other regions, where bin A reliably displayed the highest aquatic life conditions as measured by the macroinvertebrate metrics, the Ridges bin A was often characterized by slightly lower metric scores. The TN and TP concentrations that delineate this bin were very low, 0.22 mg/liter and 0.006 mg/liter, respectively. It is likely that these concentrations represent oligotrophic conditions and a bottom-up control of secondary productivity in these detritus-based upland streams. This positive response to nutrients has been previously demonstrated in empirical studies of controlled nutrient additions to headwater streams (Perrin and Richardson 1997, Cross *et al.* 2006). Metric scores were observed to generally increase through bins B and C as levels of TN increase.

The median values of TN and TP for each of the bins are presented in **Table 20** and provide an indication of overall trophic state of the streams in each bin. Generally, shifts in metric values leading to decreased scores are observed in bins D and F with higher scores in bin E. As TP concentrations were similar between bins D and E, it is suspected that degradation in bin D is more closely related to N:P ratios and resulting lowered food-quality than to effects caused by eutrophication. Bin F displayed the greatest range of biological metric values and a higher proportion of streams scoring "1" indicating that TN and TP concentrations associated with this bin may represent the beginning of nutrient assimilative impacts to the benthic community.

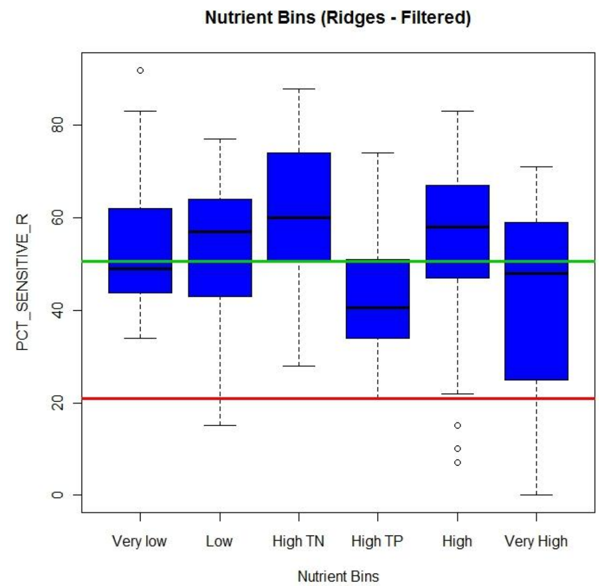


Figure 30. Percent Sensitive organisms plotted by nutrient bins in the Ridges bioregion. See **Figure 28** heading for details.

6.3.5 Quantifying nutrient degradation

In order to establish numerical criteria for nutrients in free-flowing waters, the nutrient stressor must be clearly linked to designated uses. For aquatic life, this means one must demonstrate that under specific nutrient conditions, a sample can be predicted to fail established standards for aquatic life use support a certain percentage of time. We can approximate that expectation by calculating the percentage of samples scoring "1" in each nutrient bin in a manner consistent with the formal MBSS assessment methodology. **Tables 21 - 23** provide the percentages of five selected nutrient sensitive metrics scoring "1" in each of the three tested regions. Complete results of the scoring criteria applied to the metric values can be found in **Appendix B Table 6**. It should be remembered that the physical habitat quality in these bins is not severely degraded and conductivity, dissolved oxygen, and pH are not stressful.

Table 20. Attributes of the nutrient bins presented as the median values of the data parameters in each bin.

	Coastal Plain				Piedmont/Valleys				Ridges			
BIN	TN	TP	HQI	SPCOND	TN	TP	HQI	SPCOND	TN	TP	HQI	SPCOND
A	0.25	0.012	36.5	69	0.34	0.007	45	95	0.12	0.0037	40	78
B	0.58	0.029	37.5	95	1.13	0.012	43	119	0.34	0.006	45	146
C	4.31	0.022	33	160	3.23	0.012	46	160	0.69	0.007	49	160
D	0.96	0.104	34	133	1.01	0.033	47	115	0.267	0.015	46	213
E	2.85	0.097	31	155	2.62	0.038	41.5	164	0.85	0.013	47	185
F	-	-	-	-	3.39	0.067	41	176	1.66	0.022	43	244

Table 21. Percent of samples scoring "1" in Coastal Plain nutrient bins for selected nutrient sensitive metrics.

Nutrient Bin	% Clinger	Hilsenhoff Biotic Index	% EPT	% Tolerant	% Plecoptera	Average
A	4.4%	8.7%	8.7%	4.4%	13.0%	7.8%
B	11.64%	14.0%	7.0%	17.4%	11.6%	12.3%
C	36.2%	36.2%	29.8%	55.3%	38.3%	39.2%
D	81.8%	45.5%	36.4%	63.6%	27.2%	50.9%
E	54.6%	63.6%	54.6%	72.7%	81.8%	65.5%

Table 22. Percent of samples scoring "1" in the combined Piedmont and Valleys nutrient bins for selected nutrient sensitive metrics.

Nutrient Bin	Hilsenhoff Biotic Index	% Chironomidae	% Collector	% EPT	% Tolerant	Average
A	0.0%	0.0%	5.3%	0.0%	0.0%	1.1%
B	22.9%	20.0%	18.6%	17.1%	20.0%	19.7%
C	28.7%	38.2%	14.7%	28.0%	28.0%	27.5%
D	27.3%	9.1%	18.2%	36.4%	27.3%	23.6%
E	33.3%	36.4%	24.2%	30.3%	36.4%	32.1%
F	46.2%	46.2%	46.2%	38.5%	38.5%	43.1%

Table 23. Percent of samples scoring "1" in the Ridges nutrient bins for selected nutrient sensitive metrics.

Nutrient Bin	ASPT - MOD	# EPT Taxa	% EPT	% Swimmers	# Sensitive Taxa	Average
A	22.2%	0.0%	22.2%	44.4%	11.1%	20.0%
B	8.8%	0.0%	8.8%	11.8%	5.9%	7.1%
C	9.1%	4.6%	2.3%	18.2%	4.6%	7.7%
D	50.0%	0.0%	50.0%	30.0%	0.0%	26.0%
E	12.0%	0.0%	6.0%	20.0%	4.0%	8.4%
F	61.5%	38.5%	38.5%	46.2%	38.5%	44.6%

The nutrient-sensitive metrics in bins A and B of the Coastal Plain, bins A and B of the combined Piedmont and Valleys, and bins A-E of the Ridges, shown respectively in **Tables 21 – 23**, had on average comparatively low proportions of records (≤20%) that scored “1.” The nutrient properties of these particular bins appear to be most protective of biological scores found in least-degraded or “reference” conditions. The frequency of scores of “1” doubled and tripled in the most nutrient enriched bins E and F despite the otherwise non-stressful water quality and habitat conditions.

6.4 Findings

6.4.1 Macroinvertebrate metric performance

A majority of macroinvertebrate community metrics demonstrated strong responses to nutrient conditions as expressed in the nutrient bins despite the taxonomic “coarseness” of the family-level macroinvertebrate metrics used in this study as compared to genus-level metrics. Overall, tolerance metrics and those measuring certain taxonomic groups were the best performers, followed by several habit, feeding guild, and richness metrics. Traditional tolerance metrics such as the Beck's Biotic Index, Hilsenhoff Family-Level Biotic Index, %Sensitive, %Tolerant, and EPT metrics were all around good performers. Metrics with the greatest sensitivity to the high and low nutrient bins, however, belong to the Plecoptera taxonomic group. The habit guild most responsive of nutrients was clingers across most regions; gatherers and collectors were the most sensitive feeding guilds. Additionally, the "Chessie" IBI responded negatively and significantly to increasing concentrations of TN and TP as expressed in the nutrient bins for both the Coastal Plain and the combined Piedmont and Valleys bioregions. It did not conform to patterns observed for individual metrics in the Ridges bioregion however.

Evidence for nutrient sensitivity in Plecoptera was apparent throughout this study, including initial high correlation and regression coefficients, low nutrient breakpoints observed in the recursive partitioning analyses, and finally the high frequency of poor (1) scores in the high nutrient bins across bioregions for the %Plecoptera and Plecoptera taxa count metrics. This taxonomic group is known to inhabit oligotrophic and mesotrophic bodies and has displayed overall low optimal nutrient ranges (Smith *et al.* 2007, O'Toole *et al.* 2008). Similarly, Diptera and Chironomid metrics were strong performers in the combined Piedmont and Valleys region analyses and were strongly associated with higher nutrient bins. The strong pattern of increase along the nutrient bins indicates that the presence of these generally tolerant invertebrates may be useful as indicators of nutrient enrichment.

6.4.2 MBSS BIBI and assessment methods

The filtered data set used to construct the nutrient bins turns out to be comprised almost entirely of MBSS samples as a result of the analysis requirement for water quality parameters and habitat metrics to coincide with benthic samples. In all, MBSS sampling events comprised 178 of 178 events used for the Coastal Plain, 278 of 303 Piedmont/Valley events, and 156 of 160 Ridges events in the filtered data pool. Multi-jurisdictional data set was instrumental in identifying appropriate classifications, testing individual parameters, and identifying large MDL sample sets, yet when all parameters were needed to explore and partition confounder effects, the MBSS dataset proved most useful.

Since the project is performing data analyses to support assessment of nutrient impairment in Maryland non-tidal streams, the ability of the Maryland Biological Stream Survey's Benthic Index of Biotic Integrity (BIBI) to detect nutrient degradation was investigated. MBSS BIBI scores were applied to the physiographic regions and nutrient bins developed in this study and filtered in the identical manner for confounders.

Section 6.2.6 above has already discussed the 1-3-5 scoring of the IBI metrics. For metrics which decrease with degradation, a site receives a score of 1 if it metric score is less than the 10th percentile of the scores of reference sites; a 3, if its score is between the 10th and 50th percentile of reference sites; and a 5 if its score is greater than the 50th percentile score of reference sites. (The 90th percentile is used to score a 1 if the metric increases with degradation.) An IBI greater than or equal to 3 generally means that the biological community at a site is comparable to reference sites, however, year-to-year variability in IBI scores is taken into account by calculating a minimum allowable limit (MAL), based on a comparison with the variation in biocriteria observed at MBSS sentinel sites, which are sampled annually (Southerland *et al*. 2007). The 2008 Integrated Report lists the MAL for the BIBI as 2.65.

The BIBI scores of MBSS sampling events in the filtered dataset were classified by bin for each of the three physiographic regions used in this study, just as was done for the 50 benthic metrics tested as described above. **Figure 31** shows the results. In the Coastal Plain and Piedmont and Valley regions,

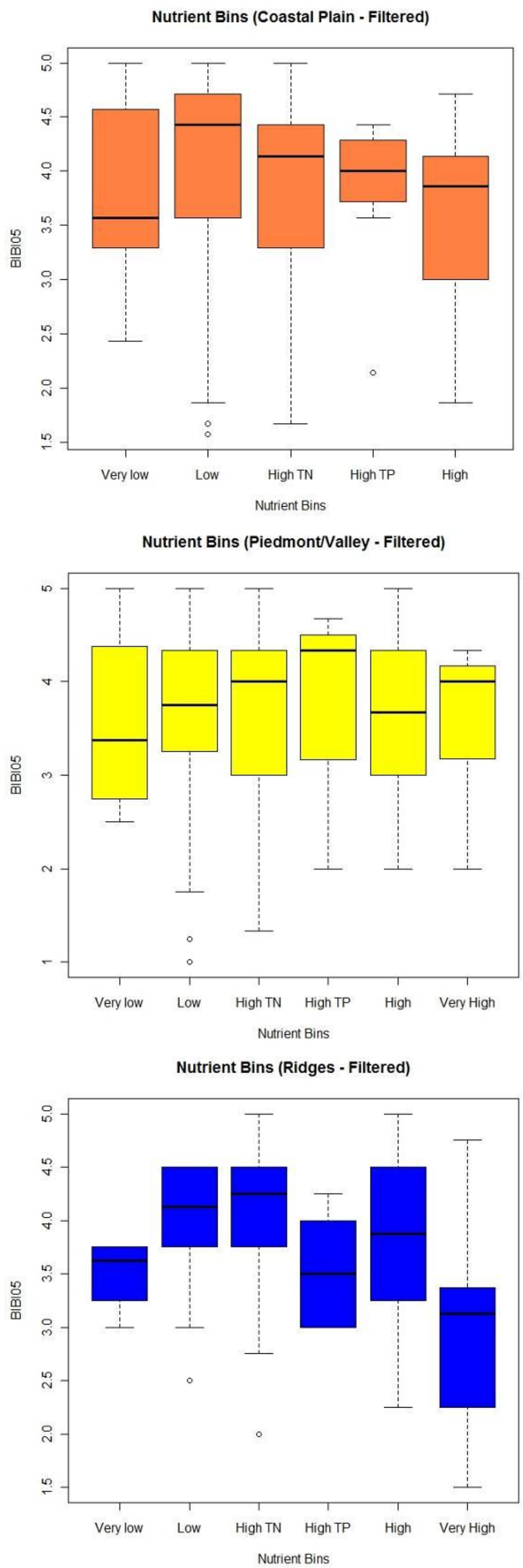


Figure 31. The MBSS regional IBIs plotted by the nutrient bins for the Coastal Plain, Piedmont/Valley, and Ridges regions.

there was no bin with a significant number of BIBI scores below the MAL of 2.65. Only in the Ridge region, in the Very High nutrient bin, were a significant number of samples below the MAL.

Four factors may contribute to the difference in response between the BIBI and the macroinvertebrate metrics tested in this study. First, MBSS uses a different classification of physiographic regions than this study. MBSS includes both the Ridge region and Valley region in a Highland region; this study, based on an exploratory data analysis, used a separate Ridge region and combined the Valley and Piedmont regions. Different classification schemes would result in different reference communities and metric scoring criteria.

Second, all of the macroinvertebrate metrics used in this study are based on family-level taxonomic identifications while some of the BIBI metrics are based on genus level identifications. Some metrics, like %EPT, are not a function of the level of taxonomic identification. For those metrics which are a function of the level of identification, Buchanan *et al*. (2011) demonstrated that family-level metrics are fully capable of discriminating reference and degraded conditions in their development of the Chessie BIBI.

More likely, differences in nutrient sensitivity can be traced to two other factors: the choice of component metrics in MBSS BIBI and differences in the definition of reference or minimally disturbed conditions. Scoring metrics is a function of the set of sites used as reference and it is not known how the set of reference sites used in the MBSS compares to the set of refined MDL sites used to score the benthic metrics in this study. **Table 24** shows the criteria used to identify MBSS reference sites. Compared to the criteria defining the refined MDL in **Table 14**, (a) the MBSS imposes more types of conditions on reference sites, but (b) where MBSS and the refined MDL impose the same type of conditions, the refined MDL tends to be more restrictive. It is not possible to tell from their definitions how MBSS reference sites and the refined MDL compare. A comparison of the metric values used to score metrics (**Table 25**), however, suggests the reference conditions are fairly similar despite differences in the selection criteria.

It should be noted that our analysis was performed on a dataset that had been filtered to remove the influence of other key stressors such as high conductivity and degraded habitat, and therefore is unlikely to contain the most degraded sites. Our analysis does not rule out the possibility that there are threshold nutrient concentrations above which the BIBI is more likely to be below the MAL. It is unlikely, though, that identification of those thresholds can be untangled from the impacts of other stressors, as has been done for the nutrient bins used in this study, and scientifically defensible nutrient thresholds are thus difficult to identify.

Table 24. MBSS reference condition criteria for streams.

pH ≥ 6 (pH<6 and DOC ≥ 8 mg/liter for blackwater streams)
ANC ≥ 50 µeq/l
DO ≥ 4 mg/l
NO_3 ≤ 4.2 mg/l
Distance to nearest road ≥ 70 m
% forested land ≥ 35%
% urban land ≤ 5%
Buffer width > 30 m
No channelization
No point sources
Instream habitat rating: optimal/suboptimal
Aesthetics rating: optimal or suboptimal

Table 25. Scoring thresholds for MBSS metrics directly comparable to those in this study.

Metrics used in MBSS Regional IBIs	Nutrient Study Thresholds 5	3	1	MBSS Scoring Thresholds 5	3	1
Coastal Plain | Coastal Plain						
Taxa Richness*	≥ 14	9 - 13	< 9	≥ 22	14 - 21	< 14
EPT Taxa Richness*	≥ 6	2 - 5	< 2	≥ 5	2 - 4	< 2
Ephemeroptera Taxa Richness*	≥ 2	0.6 - 1	< 0.6	≥ 2	1 - 1	< 1
% Intolerant Urban	-	-	-	≥ 28	10 - 27	< 10
% Ephemeroptera	≥ 17	0.6 - 16	< 0.6	≥ 11	0.8 - 10.9	< 0.8
Scraper Taxa Richness*	≥ 2	1 - 1	< 1	≥ 2	1 - 1	< 1
% Climbers	≥ 21	3 - 20	< 3	≥ 8	0.9 - 7.9	< 0.9
Piedmont/Valley | Valley						
Taxa Richness*	≥ 14	11 - 13	< 10	≥ 25	15 - 24	< 15
EPT Taxa Richness*	≥ 7	5 - 6	< 5	≥ 11	5 - 10	< 5
Ephemeroptera Taxa Richness*	≥ 3	2	< 2	≥ 4	2 - 3	< 2
% Intolerant Urban	-	-	-	≥ 51	12 - 50	< 12
% Chironomidae	≤ 4.0	4.1 - 34.5	> 34.5	≤ 4.6	4.7 - 63	> 63
% Clingers	≥ 78	43 - 77	< 42	≥ 74	31 - 73	< 31
Ridges | Highlands						
Taxa Richness*	≥ 15	12 - 14	< 11	≥ 24	15 - 23	< 15
EPT Taxa Richness*	≥ 10	7 - 9	< 6	≥ 14	8 - 13	< 8
Ephemeroptera Taxa Richness*	≥ 4	2 - 3	< 2	≥ 5	3 - 4	< 3
% Intolerant Urban	-	-	-	≥ 80	38 - 79	< 38
% Tanytarsini	-	-	-	≥ 4	0.1 - 3.9	< 0.1
% Scrapers	≥ 11	3 - 10	< 2	≥ 13	3 - 12	< 3
% Swimmers	≥ 12.5	2 - 12.4	< 2	≥ 18	3 - 17	< 3
% Diptera	≤ 16	17 - 41	> 41.1	≤ 26	27 - 49	> 50

7. Discussion and Recommendations

7.1 Summary of Findings

The available monitoring data were explored for evidence of phytoplankton, periphyton, and macroinvertebrate responses to the nutrients nitrogen and phosphorus in Mid-Atlantic non-tidal streams and rivers. Some natural variability was reduced by classifying biological response metrics by physiographic region and stream size class. Several non-nutrient environmental factors obscured or confounded the responses. Nutrient responses became apparent only after a) they were considered in the context of the light environment, or b) data records associated with stressful levels of the confounding environmental factors were removed from the analysis. Nutrient responses of river phytoplankton were not evident unless the water clarity surrogates turbidity and dissolved organic carbon were simultaneously considered. Periphyton nutrient responses in 1^{st} – 4^{th} order streams were evident only after records associated with marginal/poor stream bank conditions and high conductivity were removed. Macroinvertebrate nutrient responses in 1^{st} – 4^{th} order streams were evident only after records associated with high conductivity and marginal/poor in-stream habitat quality, and to a less degree extreme pH levels and low dissolved oxygen, were removed.

Biological metric responses were tested against an abiotic condition gradient consisting of 5 - 8 distinct nitrogen and phosphorus categories, or “bins.” The phytoplankton bins spanned the entire range of observed conditions. The macroinvertebrate and periphyton bins spanned a shorter range of conditions because data records with severely degraded physical habitats and stressful chemistry were removed. The boundaries of each bin were created from TN and TP breakpoints identified in the data with recursive partitioning. Concentrations in the bins ranged from “very low” TN and TP to “very high” TN and TP, and included intermediate categories such as “high TP & low TN” and “mixed” (where bin boundaries were set by light attenuation levels rather TP and TN). Nitrogen and phosphorus concentrations varied within the boundaries of each bin and their effects on biological metrics were considered in the context of each other and—for phytoplankton—in the context of the light environment.

7.1.1 Protective thresholds

Thresholds protective of high quality biological communities are apparent in the results. Thresholds are considered protective of high quality if the biological response metric meets desirable endpoints in all or most instances. In Section 4.4 above, we identified desirable phytoplankton endpoints as few (<2%) chlorophyll *a* concentrations above 30 µg/liter (“algal bloom”) and a median concentration less than 3.2 µg/liter. A desirable endpoint for algal periphyton is tentatively identified as a chlorophyll *a* concentration below 100 mg/m^2 based on Stevenson and Bahls (1999). For macroinvertebrates, we developed a metric scoring approach similar to the one employed by MBSS, where a score of “1” indicates a status most unlike that observed in an independently defined high quality environment (Section 6.3.5). The desired endpoint for macroinvertebrate metrics developed with this process was, for five bioregion-specific nutrient-sensitive metrics, an average score of “1” (poor) in 20% or fewer samples; or conversely an average score of “3” or “5” in greater than 80% of samples.

Table 26 summarizes the median concentrations of the nutrient (and co-varying light attenuation) parameters in the most nutrient enriched bins that still met the desirable endpoints. We propose these conditions as thresholds that protect against nutrient-related degradation in higher quality Maryland

Table 26. Environmental conditions protective of high quality biological communities. Listed are the median concentrations of total phosphorus and total nitrogen (and the light co-variables turbidity and DOC for phytoplankton) in the most nutrient enriched data analysis bins associated with high quality biological communities, and the conditional levels of the identified confounding parameters. Biological quality begins to decline when nutrient (and co-variable) concentrations increase above these levels and/or one or more confounding factors fails the indicated thresholds. High quality biological communities: phytoplankton median chlorophyll *a* concentration less than 3.2 µg/liter and frequency of algal blooms (i.e., >30 µg/liter) less than 2%; nuisance levels of algal periphyton chlorophyll *a* concentration (i.e., >100 mg/m^2) absent; the five selected nutrient-sensitive macroinvertebrate metrics in each physiographic region score "1" (poor) in 20% or fewer samples.

Response Variable	Phytoplankton Chl *a*						Periphyton Chl *a*	Macroinvertebrate Family-Level Metrics		
Physiographic Region(s)	Piedmont, Ridges, Valleys		Mid-Atlantic Coastal Plain (MACP)		Southeastern Plain (SEP)		Piedmont, Ridges, Valleys	Piedmont, Valleys	Ridges	Coastal Plain (MACP & SEP)
Strahler Stream Order	$5^{th} – 7^{th}$		$1^{st} – 5^{th}$		$1^{st} – 5^{th}$		$1^{st} – 4^{th}$	$1^{st} – 4^{th}$	$1^{st} – 4^{th}$	$1^{st} – 4^{th}$
Analysis Bin	4	4	2	2	3+4	3+4	5	B	E	B
DOC level	Lo	Hi	Lo	Hi	Lo	Hi	n/a	n/a	n/a	n/a
<u>Nutrients and Light Co-Variants</u>										
Total phosphorus (mg/liter)	0.036	0.087	0.012	0.030	0.059	0.085	0.050 [1]	0.012	0.013	0.029
Total nitrogen (mg/liter)	2.44	2.37	2.36	2.15	2.67	1.19	0.93 [1]	1.13	0.85	0.58
Turbidity (NTU)	10.0	10.0	5.0	4.7	6.3	8.9				
DOC (mg/liter)	2.16	3.81	2.37	4.73	2.37	4.85				
<u>Conditional Requirements at Site for Nutrient Responses to be Evident</u>								Piedmont ǀ Valleys		
Spec. Conductivity (µS)							?	<200 ǀ <329	<282	<200
pH								6 - 9	5.5 - 9	6 - 9
Diss. Oxygen (mg/liter)								>5.0	>5.0	>5.0
Habitat Quality Index [2] /60								≥30 ǀ ≥30	≥35	≥25
Channel Alteration Score /20							>10			
Bank Stability Score /20							>10			
Riparian Vegetation Score /20							>10			
Bank Vegetation Score /20							>10			
Karst bedrock							not present			

[1] one year average of nutrient concentrations at neighboring sampling locations (see 5.2 for details)
[2] sum of scores for embeddedness, epifaunal substrate, and riffle frequency/quality, with a possible range of 0 - 60

stream and river biological communities. Nutrient concentrations above these median values will increase the frequency of undesirable endpoints in phytoplankton, algal periphyton, and macroinvertebrates.

Protective thresholds for water column TP range from 0.012 to 0.087 mg/liter. Phytoplankton thresholds in the Mid-Atlantic Plain are somewhat lower than those in the Southeastern Plain and the combined Piedmont, Ridges, and Valleys region, and evidence suggests this is related to regional differences in water chemistry and geology. Note that phytoplankton communities in high DOC environments may be able to maintain desirable Chla levels at higher TP concentrations but require slightly lower TN concentrations. The TP threshold protective against nuisance levels of algal periphyton (0.05 mg/liter) in the combined Piedmont, Ridges, and Valleys streams is tentative because of the relatively small size of the analysis data set. However, it falls mid-way in the overall range of protective thresholds. TP thresholds protective of macroinvertebrate endpoints in streams are lower than those developed for periphyton and phytoplankton, ranging between 0.012 and 0.029 mg/liter. This may be a function of the very nutrient-sensitive metrics selected to establish a desirable endpoint. It may also be a function of higher phosphorus concentrations in macroinvertebrate food sources. Many macroinvertebrate taxa feed on periphyton and detritus. Periphyton have the ability to actively take up phosphorus from the water column and store it, so phosphorus concentrations in the water column are typically lower than those in the periphyton biomass. Phosphorus also adsorbs to detritus and is used by associated microfauna.

Protective thresholds for water column TN range from 0.58 to 2.67 mg/liter. Thresholds protective of phytoplankton endpoints are again higher than those protective of periphyton and macroinvertebrate endpoints. Comparisons of the nutrient concentrations and macroinvertebrate responses in analysis bins C and D are interesting to consider. Bin C experiences TP concentrations that are comparable to or lower than those in bin A (Very Low) and B (Low) but has roughly 2X to 7.4X higher TN concentrations. Bin D experiences the opposite, with 2.5X to 4X higher TP and TN levels comparable to those in bins A and B (**Table 20**). The nitrogen increase in bin C led to an 8% increase in “poor” scores in Piedmont/Valleys and a 27% increase in “poor” scores in the Coastal Plain (**Tables 21-22**). The strong negative response to nitrogen confirms the potential of excess nitrogen to stress biological communities in these bioregions, and protective TN thresholds may assist in maintaining high quality biological communities. In the nutrient-limited circumstances of the Ridges bioregion, biometric responses are different. The nitrogen increase expressed in bin C did not elicit a strong response whereas the phosphorus increase in bin D elicited an 18% increase in “poor” scores. This suggests that either a) the increases in both nutrients did not exceed limitation thresholds, or b) the low phosphorus levels in the Ridges’ bin C are holding back any biological responses to nitrogen increases. The periphyton bins 3 and 4 in the combined Piedmont, Ridges, and Valleys region create the same contrasting nutrient environments but with 1 year averages of TN and TP (**Table 12**). Periphyton chlorophyll *a* content and phosphorus content increase significantly in response to the nitrogen increase expressed in bin 3. These results are interesting in light of the work of Vollenweider (Vollenweider *et al.* 1980) and others that points out the primary importance of phosphorus loadings to lake eutrophication. Mid-Atlantic streams appear to respond to both nitrogen and phosphorus increases when concentrations exceed the protective thresholds above.

The Ridges bioregion offers an example of how the characteristics of a particular data set can shape the RPART and binning results. Most of the filtered records in the Ridges data set had low and moderate TN and TP levels; many of the high TN and TP records were confounded by non-nutrient factors and removed. Multiple nutrient breakpoints were still identified in the RPART models run on the Ridges’ filtered data. They were successfully used to create a series of distinct nutrient bins, most with low TN and/or TP concentrations, and significant nutrient-related degradation in the macroinvertebrate community appeared in the "High TP & Low TN" bin (D) and "Very High" bin (F). Despite differences in the underlying data sets, there is good agreement across bioregions as to what nutrient levels correspond to desirable biological endpoints. Specifically, TP and TN concentrations in the Ridges' bins D and F are roughly equivalent to those in the "Low" bins (B) of the Piedmont/Valleys and Coastal Plain. The RPART method seems to be able to indicate meaningful nutrient breakpoints for identifying protective thresholds as long as there is a sufficient range of nutrient conditions.

The overall range of protective thresholds of 0.012 - 0.087 mg TP/liter and 0.58 - 2.67 mg TN/liter identified in this study overlap similar thresholds identified in other studies. Applying a quantile analyses technique to the MBSS database, Morgan and Kline (2010) identified an upper stream TP criterion between 0.025 and 0.037 mg/liter and an upper stream TN criterion between 1.34 and 1.68 mg/liter as protective of small stream integrity in Maryland. This study's range of protective thresholds overlaps the 0.039 – 0.064 mg/liter range of TP endpoints identified by Paul and Zheng (2007) for Maryland wadeable streams, the TN endpoint of 1.8 mg/liter identified by Zheng *et al.* (2008) for West Virginia wadeable streams, and the 0.07 mg TP/liter and 2.01 mg TN/liter endpoints identified by Sheeder and Evans (2004) for Pennsylvania watersheds. The TP and TN thresholds also agree well with endpoints developed by researchers outside the Maryland river basins, including Robertson *et al.* (2006, 2008) and Weigel and Robertson (2007) for Wisconsin, and Smith *et al.* (2007) for New York. Except for the recent Morgan and Kline study, these other studies are discussed in more detail in ICPRB's nutrient criteria literature review (Mandel *et al.* 2010).

7.1.2 Confounding environmental parameters

Non-nutrient stressors—particularly turbidity, conductivity, and physical habitat—were repeatedly selected by the RPART application over TN and TP in the unfiltered data. To identify nutrient-related degradation in each analysis data set, the effects of the non-nutrient factors had to be filtered from the data set or minimized to the greatest extent possible. This filtering and minimizing effectively created a data subset with generally unstressed, "good" environmental conditions with which to find nutrient breakpoints and thresholds. Taken out of the context of these good environmental conditions, nutrient breakpoints and thresholds are not easily discernable because of the stronger impacts of other parameters, even though nutrient impacts may be occurring.

Development of nutrient criteria will need to consider nutrient effects in the context of other, non-nutrient stressors. Nutrient reductions to meet numerical nutrient criteria may not lead to improvement of aquatic life condition if other stressors are present at levels harmful to aquatic life uses. There was some analytical evidence for lower nutrient breakpoints in streams with poor habitat quality, indicating that aquatic life already stressed by significantly degraded environments may be unable to assimilate and metabolize nutrients as well as stream communities in intact environments. In the macroinvertebrate RPART trees, there were consistently lower nutrient breakpoints in the poorer habitat sides of the splits. These results, however, would need to be explored further to be demonstrated more conclusively.

It is evident from the nutrient binning results that phytoplankton, periphyton and macroinvertebrates are responding to nutrient gradients rather than to a single threshold of degradation or impairment. In fact, multiple thresholds exist in different environmental circumstances, and could be used as criteria depending on the level of protection intended for a particular stream segment. As discussed in the next section, thresholds for impairment of a designated aquatic life use are difficult to identify because they are predicated on policy decisions that quantify how much failure is acceptable.

7.2 Applicable Water Quality Standards

7.2.1 Stressor Response Relations and Designated Uses

The purpose of numerical water quality criteria is to identify the appropriate range of the concentration of a substance that supports a designated use. When excess amounts of a substance threaten a designated use, which is the presumption for nutrients, numerical criteria usually take the form of concentration thresholds marking an increased risk of failing a designated use.

The Scientific Advisory Board (SAB), which reviewed the stressor-response methodology and the draft USEPA guidance document, identified the link between stressor-response relationships and designated use support as a major weakness of the USEPA's formulation of the stressor-response methodology. SAB (2010) urged the EPA to address the following issues:

> 1) establish linkages among the designated uses and measured responses, stressor and measures of stressors; and
> 2) relate measures of stressors directly to deleterious effects on designated uses.

For each stressor-response relationship determined in this analysis, **Table 27** shows the relevant designated uses and the methods used to determine whether or not they are supported. The existence of a method for determining use support is a necessary condition for relating a quantitative stressor-response relationship to deleterious effects on designated uses. As **Table 27** shows, not all designated uses have methods for determining use support.

The support of aquatic life is generally considered the primary designated use threatened by excess nutrients. Maryland has standardized its methodology for determining aquatic life use impairment in small (1^{st} - 4^{th}) order streams, but not in larger rivers.

Excess phytoplankton can cause taste and odor problems in drinking water, or interfere with water treatment, so the use of a waterbody for drinking water supply can also be impaired if excess nutrients lead to excess phytoplankton growth. Maryland has developed water quality criteria for drinking water reservoirs, but not for free-flowing rivers and streams. (In this analysis we have borrowed the drinking water reservoir criteria for chlorophyll *a* and applied it to free-flowing waters.)

Both excess phytoplankton and excess periphyton can be unseemly and in that sense diminish the enjoyment of primary and secondary contact recreation, but aesthetic criteria are secondary issues that are less amenable to treatment by quantitative criteria. They will not be considered further in this analysis.

The relevant designated uses for each stressor-response relationship will be discussed in more detail below. Based on the current standards, only the relationship between benthic macroinvertebrates and nutrients can be directly related to use support. For the other cases, recommendations will be given for how to close the gap between the stressor-response relationship and support of the designated use. These recommendations will be explained in more detail below. **Table 27** briefly summarizes the recommendations and identifies the sections where they are discussed.

7.2.2 Phytoplankton and the Water Supply Use

The threshold concentration for algal blooms, 30 µg/liter, was taken from Maryland's proposed chlorophyll *a* water quality criteria for reservoirs (MD 2010b). The criteria have two components:

1. the 90th percentile concentration of observed chlorophyll *a* should be no greater than 30 µg/liter; and
2. the 30-day moving average chlorophyll *a* concentration should be no greater than 10 µg/liter.

The two components work in concert. Chlorophyll *a* concentrations above 30 µg/liter are associated with blooms or nuisance-level algae populations which are marked by a shift to the dominance of cyanobacteria (blue-green algae). These blooms can produce taste and odor problems or other water treatment problems such as increased trihalomethane precursors (Walker 1984). An average chlorophyll *a* concentration of 10 µg/liter represents the boundary between mesotrophic and eutrophic conditions in lakes (Carlson 1977). Using existing data sets from U. S. Corps of Engineer reservoirs as well as reservoirs in South Africa and Vermont, Walker (1984) developed a regression model which predicts the frequency of instantaneous chlorophyll *a* concentrations over 30 µg/liter as a function of average chlorophyll *a* concentrations, and found that the frequency increased when the average concentration was greater than 10 µg/liter.

It has not been established, however, that these relations hold in free-flowing rivers. We have identified two datasets that could be analyzed to confirm if the same relationships occur in other data sets for free-flowing rivers: the Ohio River Valley Water Sanitation Commission (ORSANCO) data set for the Ohio River basin and the Washington Aqueduct data set for the Potomac River mainstem near the fall-line. We recommend pursuing this analysis. Another line of analysis that could be pursued is relating nutrient or chlorophyll *a* concentrations to geosmin or 2-MIB, cyanobacteria metabolism by-products which have been implicated in taste and odor problems in drinking water. Dzialowski *et al.* (2009) were able to develop reservoir-specific empirical regression models for geosmin for five water supply reservoirs in Kansas. Total algal biovolume, total cyanobacterial biovolume, and chlorophyll *a* were

Table 27. Designated use and Maryland impairment determination method currently in place.

Taxonomic Group Related to Nutrient Stressors	Designated Use	Impairment Determination Method	Recommendations for Next Steps to Link Stressor to Designated Use	Details in Section
Phytoplankton	Water Supply	None	Analyze river chlorophyll *a* data to relate to species shifts and treatment problems	7.2.2
	Aquatic Life (Large Rivers)	None	Develop biological assessment methodology	7.2.3
	Aquatic Life (Coastal Plain)	Biocriteria Assessment Based on MBSS Indices of Biological Integrity	Relate chlorophyll *a* concentrations to benthic and fish metric scores	7.2.3
Periphyton	Aquatic Life	Biocriteria Assessment Based on MBSS Indices of Biological Integrity	Relate periphyton biomass to benthic and fish metric scores	7.2.3
Benthic Macroinvertebrates	Aquatic Life	Biocriteria Assessment Based on MBSS Indices of Biological Integrity	Develop Tiered Aquatic Life Uses and Biological Condition Gradient	7.2.4

among the predictive variables used in their models. Dzialowski *et al.* (2009) found, however, that geosmin concentrations were negatively related to these variables. They speculate that their results were due to the fact that they did not measure and distinguish dissolved geosmin from its cell-bound or particulate forms. In particular, they note that the release of geosmin in reservoirs may be triggered by the on-set of phosphorus limitation to algal growth, in contrast to rivers, where nitrogen limitation may induce geosmin release from cyanobacteria algal mats.

The Potomac River is the major source of water supply for the Washington metropolitan area. According to the Washington Suburban Sanitary Commission and the Washington Aqueduct, two of the major drinking water suppliers in the area, taste and odor problems have not been associated directly with water taken from the Potomac River (Chen 2011, Stowe 2011). The Washington Aqueduct sometimes has had taste and odor problems associated with algal growth in two reservoirs used to hold water before treatment (Stowe 2011). The water utilities do find, however, that under low-flow conditions and particularly in the summer, the water is more difficult to treat. It is more difficult to coagulate solids and more pre-oxidants must be added for effective treatment. The utilities associate these difficulties with increased algal activity. The Washington Aqueduct has collected algal taxa counts, chlorophyll *a* concentrations, and nutrient concentrations from the Potomac River but has yet to find a meaningful relation between measures of algal activity and treatment problems. This dataset is one of the two datasets we recommend using above for further analysis of the relation between nutrients, chlorophyll *a*, and the use of river for water supply.

7.2.3 Phytoplankton, Periphyton, and Aquatic Life Use Impairment

Maryland's water quality standards do not have chlorophyll *a* criteria for periphyton, to protect aquatic life in free-flowing streams and rivers. Aquatic life use support in small order streams is measured by fish and benthic macroinvertebrate indices of biological integrity (FIBI, BIBI). The relationship between chlorophyll *a* concentrations, or other measures of algal biomass, and the fish and macroinvertebrate metrics comprising the IBIs is the missing link connecting nutrients to aquatic life use support.

There is insufficient periphyton monitoring data in Maryland to determine that link. We recommend that additional monitoring data be collected. We also propose that a pilot analysis relating chlorophyll *a* or other measures of periphyton biomass to fish and benthic metrics could be performed on periphyton data collected outside of Maryland.

For streams and rivers in the Coastal Plain, it may also be possible to determine a relationship between observed chlorophyll *a* concentrations and macroinvertebrate metrics. One major hurdle to such an analysis is that the periphyton and phytoplankton chlorophyll *a* data are not typically collected at the same locations or at the same time as the fish and benthic monitoring data, so any analysis relating benthic macroinvertebrate metrics through periphyton and phytoplankton directly back to nutrient concentrations needs to be performed with temporally and spatially separated data. We recommend performing an exploratory analysis which first correlates chlorophyll *a* monitoring data with benthic sampling locations by applying the approach used in Section 5 of this study which related VADEQ ambient monitoring stations to macroinvertebrate monitoring locations. If a reasonable relation between monitoring locations can be established, a preliminary analysis of the relationship between chlorophyll *a* concentrations and biological metrics could be performed that could help determine if it would be worthwhile to collect chlorophyll *a* data concurrently with biological monitoring data to put the analysis on a firm footing.

As mentioned above, Maryland does not have an established methodology for determining aquatic life use impairments in rivers larger than 4th order. Establishing clear-cut criteria for determining the impairment or support of aquatic life use in large rivers is a prerequisite for determining a nutrient or chlorophyll *a* criteria to protect that use.

7.2.4 Aquatic Life Use Impairment in Streams and the MBSS Benthic Index of Biotic Integrity

The primary biocriteria used in Maryland's assessment of aquatic life use support are two multi-metric indices of biological integrity, one for fish (FIBI) and one for benthic macroinvertebrates (BIBI), which are based on MBSS biological monitoring data (Roth *et al*. 2000; Stribling *et al.* 1998). For each sample site, the IBI's are calculated from component metrics scored on thresholds identified from reference or least-degraded sites, i.e. sites with minimal human impact as defined by the MBSS. Southerland *et al*. (2005) describe rationale for the selection of metrics.

Nutrient criteria are intended to identify nutrient concentration thresholds above which impairment to aquatic life is likely to occur, taking into account the confounding influence of other stressors. In terms of the FIBI or BIBI used in Maryland to measure aquatic life support, nutrient concentrations above the thresholds should make it more likely that a site receives an IBI score less than 3, or more specifically, less than the MAL, or minimum allowable limit. Our analysis has demonstrated that a wide variety of macroinvertebrate metrics are sensitive to nutrient concentrations once confounding factors are accounted for or removed. We have identified potential protective nutrient criteria based on the total nitrogen and total phosphorus concentrations in bins that meet desirable biological endpoints. These bins appear to represent ecologically significant conditions. For the bins to be used to develop nutrient criteria in Maryland, a significant number of the sites in one or more of the nutrient bins should have IBI scores below the MAL.

BIBI scores calculated by the MBSS program were classified into the macroinvertebrate nutrient bins for each of the three physiographic regions used in this study, just as was done for the 50 macroinvertebrate test metrics. The records were filtered to remove those confounded by other environmental factors. **Figure 31** shows the results. In the Coastal Plain and Piedmont and Valley Regions, there was no bin with a significant number of MBSS BIBI scores below the MAL of 2.65, even in bins with high nutrient concentrations. Only in the Ridge Region, in the Very High nutrient bin, were a significant number of samples below the MAL.

What explains the relative insensitivity of the BIBI to the gradient of nutrient bins, in contrast to many of the other macroinvertebrate metrics discussed in Section 6? The difference in the set of sampling events analyzed is minimal. Although the multi-jurisdictional data set was instrumental in identifying appropriate classifications, testing individual parameters, and identifying large MDL sample sets, the MBSS dataset proved most useful when all parameters were needed to explore and partition confounder effects. The filtered data set used to construct the nutrient bins turns out to be comprised almost entirely of MBSS samples as a result of the analysis requirements for water quality parameters and habitat metrics to coincide with benthic samples. In all, MBSS sampling events comprised 178 of 178 events used for the Coastal Plain, 278 of 303 Piedmont/Valley events, and 156 of 160 Ridges events in the filtered data set.

Section 6.4.2 presents and discusses four possible reasons why the MBSS BIBI did not show nutrient responses:

1) MBSS uses a different classification of physiographic regions than this study

2) All macroinvertebrate metrics used in this study are based on family level taxonomic identifications, while some metrics in the MBSS BIBI are based on genus level identifications (this affects metrics like taxa richness but not ones like %EPT)
3) Some of the metrics comprising the MBSS BIBI were found to be nutrient insensitive in this study, which would tend to dampen the overall BIBI sensitivity to nutrients
4) The reference conditions used to establish the metric scoring criteria are defined slightly differently by MBSS (**Table 25**) and this study (**Table 14**)

The component metrics of the MBSS BIBI were selected for their ability to differentiate reference conditions from degraded conditions as quantified by habitat and water chemistry, not for their sensitivity to nutrients. Results of the analysis above suggest that some of the metrics used in the MBSS BIBI are not responsive to nutrient concentrations. Our approach of removing from the analysis data set those records significantly impacted by other environmental factors is effectively capturing nutrient impacts in high quality waters as well as waters of lesser quality that still support their aquatic life uses. We recommend a closer comparison of the refined "minimally disturbed locations" used in this study and the reference sites used to establish scoring thresholds for metrics in the MBSS BIBI. The USEPA has been encouraging states to recognize in their standards that there are levels or tiers of biological health or integrity. The impacts of nutrients on aquatic life may best be explained within the context of these tiered aquatic life uses (TALUs). This is discussed in more detail in the following section.

7.3 Tiered Aquatic Life Uses and the Biological Condition Gradient

We believe the results of this study can best be interpreted in the context of the Tiered Aquatic Life Uses (TALUs) concept and the associated Biological Condition Gradient (BCG) concept. TALUs use biological information to define a set of Aquatic Life Uses for waterbodies on a scale relative to their "natural conditions" (USEPA 2005). They are an attempt to bring the protection of the aquatic environment a step closer to the overarching goal of the Clean Water Act (CWA), which is to "restore and maintain the chemical, physical, and biological integrity of the Nation's waters." TALUs can be explained by invoking the concept of a biological condition gradient (BCG), although TALU and BCG are independent concepts.

7.3.1 Tiered Aquatic Life Uses (TALU)

Biological integrity can be defined as "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region" (Karr and Dudley 1981). In 1977, at the time of the inception of the CWA, restoring and maintaining biological integrity was a worthy, but imprecise, goal. The CWA set as an interim goal "the protection and propagation of fish, shellfish, and wildlife." Karr and Dudley (1981) suggest that the interim goal is primarily concerned with achieving the water quality, i.e. the physical and chemical conditions, necessary for the protection and propagation of aquatic life. For that reason, the goals of the CWA are expressed in terms of water quality standards which identify the designated uses of the waterbody and water quality criteria necessary to support those uses. With the development and widespread use of biocriteria for water quality assessment, however, the opportunity exists to define designated uses directly in terms of biological measures that can directly represent the goal of biological integrity.

In addition to determining designated uses for supporting aquatic life in terms of biological metrics, and not just the physical and chemical conditions that make them possible, TALUs also incorporate the concept that a single designated use to support aquatic life may not be appropriate for all waterbodies. Not only can waterbodies differ in respect to their natural condition, which may vary as a function of size, geology, and other factors, but restoring the natural condition of a waterbody may not be the appropriate goal. This recognition leads to the concept of a scale or tiers of aquatic life uses. The Federal Code of Regulations allows for states to "…adopt sub-categories of a use and set the appropriate criteria to reflect varying needs of such sub-categories of use… (40CFE131.10(c))" Several states, including Ohio, Maine, and Texas, have adopted tiered aquatic life uses which give greater protection to high quality waters and set lower standards for waters for which restoring natural conditions is not judge to be attainable. Davies and Jackson (2006) list the following benefits from adopting tiered aquatic life uses in Maine:

- Facilitates identifying and preserving high quality waters;
- More accurately represents current conditions;
- Helps sets attainable management goals;
- Preserves incremental improvements; and
- Triggers management action if biological conditions decline.

7.3.2 Biological Condition Gradient (BCG)

BCG is a conceptual model or framework of how aquatic ecosystems respond to increasing anthropogenic disturbances. BCG begins with the use of undisturbed natural conditions as a reference point to define the state of aquatic systems. Aquatic ecosystems deviate from those natural conditions in response to increasing human disturbance and the accompanying increase in the level of physical, chemical, and biological stressors. **Figure 32** illustrates the gradient in biological conditions in response to the gradient of stressors and human disturbance.

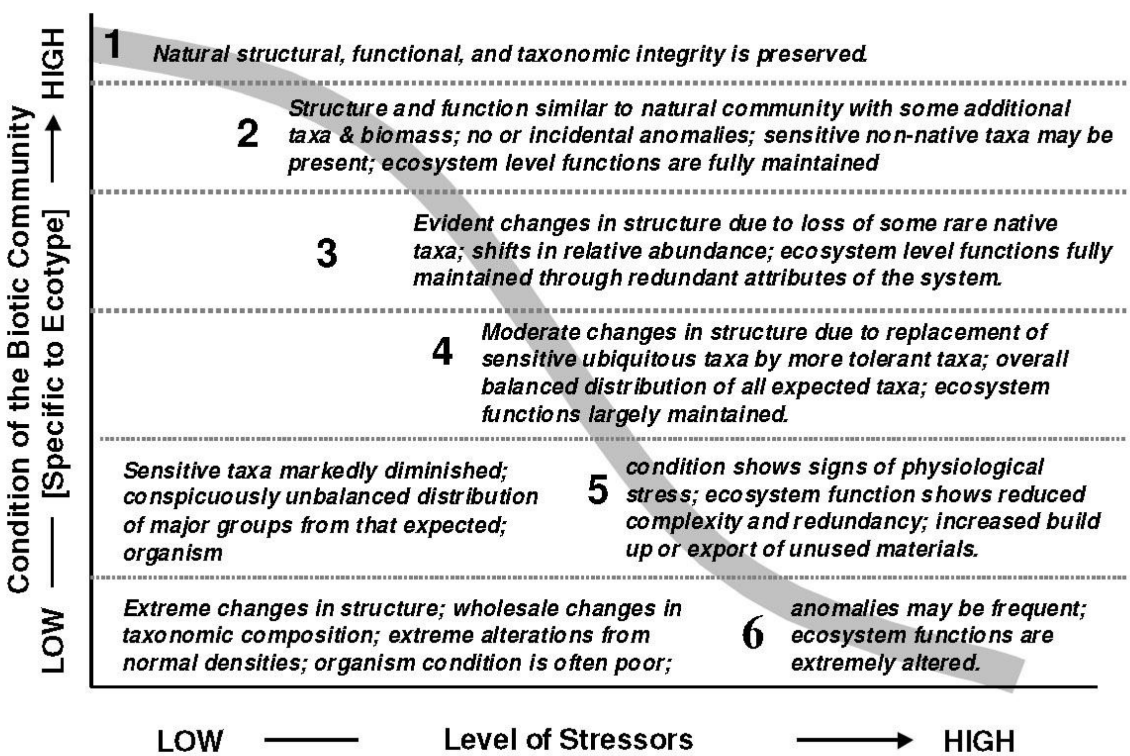


Figure 32. Conceptual model of the Biological Condition Gradient (BCG) adapted from USEPA (2005).

Under the BCG framework, the gradient of biological conditions, from natural conditions to the most degraded conditions, can be described in six tiers, according to the taxonomic, structural, and functional characteristics of the aquatic ecosystem. The six tiers are more fully specified by ten ecological attributes, given in **Table 28**. A full characterization of the six tiers in terms of the ten ecological attributes forms the Biological Condition Gradient Matrix. **Table 28** also gives the characterization of Tier 4, the lowest tier considered to satisfy the requirements of the aquatic life use support, in terms of

Table 28. Ecological attributes of Tier 1 and 4 of the Biological Condition Gradient (USEPA 2005). Tier 1 is the natural, unaltered condition. Tier 4 is the lowest tier considered to satisfy the requirements of the aquatic life use support.

Ecological Attribute	Tier 1	Tier 4
I Historically documented, sensitive, long-lived or regionally endemic taxa	As predicted for natural occurrence except for global extinctions	Some may be absent due to global, regional or local extirpation
II Sensitive-rare taxa	As predicted for natural occurrence, with at most minor changes from natural densities	May be markedly diminished
III Sensitive-ubiquitous taxa	As predicted for natural occurrence, with at most minor changes from natural densities	Present with reproducing populations maintained; some replacement by functionally equivalent taxa of intermediate tolerance.
IV Taxa of intermediate tolerance	As predicted for natural occurrence, with at most minor changes from natural densities	Common and often abundant; relative abundance may be greater than Sensitive-ubiquitous taxa
V Tolerant taxa	As naturally occur, with at most minor changes from natural densities	May be common but do not exhibit significant dominance
VI Non-native or intentionally introduced taxa	Non-native taxa, if present, do not displace native taxa or alter native structural or functional integrity	Some replacement of sensitive nonnative taxa with functionally diverse assemblage of non-native taxa of intermediate tolerance
VII Organism Condition (especially of long-lived organisms)	Any anomalies are consistent with naturally occurring incidence and characteristics	Incidence of anomalies may be slightly higher than expected
VIII Ecosystem Functions	All are maintained within the natural range of variability	Virtually all are maintained through functionally redundant system attributes though there is evidence of loss of efficiency (e.g., increased export or decreased import)
IX Spatial and temporal extent of detrimental effects	N/A A natural disturbance regime is maintained	Mild detrimental effects may be detectable beyond the reach scale and may include more than one season
X Ecosystem connectance	System is highly connected in space and time, at least annually	Some loss of connectance but colonization sources and refugia exist within the catchment

the ten ecosystem attributes. USEPA (2005) provides the full Biological Condition Gradient Matrix for all tiers and attributes.

Forty-four aquatic ecologists from 23 states took part in an evaluation of the BCG model (Davies and Jackson 2006) to test how consistently individual biologists would use the BCG to classify biological monitoring data. Thirty-three scientists worked on classifying 54 benthic samples, while 11 scientists classified 58 fish samples. The samples came from six broad geographic regions across the county. The scientists classifying benthic samples agreed on which tier to assign a sample 82% of the time and the scientists classifying the fish samples agreed 74% of the time. In most cases classifications diverged by no more than one tier.

One of the goals of BCG is to provide a framework for comparing different biological monitoring and assessment programs. States and territories used different biocriteria to assess their waters. By interpreting those biocriteria in terms of BCG, the programs can be compared to each other. The BCG approach also provides a framework for determining the consistency of biocriteria. It can help improve communication of biomonitoring results and implications among scientists, managers, and the public.

7.3.3 Generalized Stressor Gradient (GSG)

The Generalized Stressor Gradient (GSG) is the complement to the BCG. It is the x-axis in **Figure 32**, and it measures and scales disturbances which impact the biological community. The GSG should include such factors as flow regime, energy sources, habitat structure, conventional and toxic chemical factors, and biotic factors such as invasive species and disease. Excess nutrients are among the factors that contribute to the stresses on a biological community.

EPA (2005) lays out a step-by-step methodology for developing a regional BCG. The first step is to characterize the natural conditions appropriate for a waterbody taking into account differences in stream size, geology, and other factors that might distinguish natural aquatic communities. It is possible that no existing sites represent natural conditions and the determination of natural conditions will have to rely on historical description. The next step is to qualitatively identify potential stressors and the aquatic communities likely to emerge under the influence of those stressors. This is in effect to provide a regional specification of the Biological Condition Gradient Matrix.

Assuming that the biological monitoring program provides sufficient information, the final step is to develop quantitative rules to define the tiers by calibration. In this step, existing biological data are first assigned to each tier, based on monitored characteristics of the biota and professional judgment. Biological taxa are classified according to how rare or common they are, how sensitive or tolerant they are to stressors, and whether they are native to a region, introduced, or invasive. This provides the foundation for quantifying the first six ecological attributes that characterize the tiers. Finally, thresholds for biological metrics are determined, based on their ability to predict the classification of the biological samples previously assigned to tiers by consensus of aquatic ecologists. Gerritsen and Leppo (2005) document the application of this methodology to develop BCGs for New Jersey streams.

If a GSG could be developed concurrently with a BCG, their relationship would resemble the traditional dose-relationship of toxicological studies. Like the mortality induced by a dose a toxic chemical, a level of stress could be related to degree of degradation of a biological community. Unlike the BCG, however, where there are clear-cut examples of its application, the development of the GSG remains theoretical. EPA (2005) concedes that a single stressor gradient may not be appropriate way to represent the impact of the multiple factors influencing aquatic health.

7.3.4 Characterizing environmental stressors

The conductivity and habitat quality breakpoints identified in this study provide an indication of when changes in these two parameters begin to affect macroinvertebrate and periphyton communities and mask nutrient and other responses. Turbidity and DOC breakpoints are useful for classifying streams into water quality categories more or less susceptible to phytoplankton algal blooms. The nutrient breakpoints identify when, in a minimally disturbed environment, TN and/or TP concentrations begin to degrade stream and river biological communities. This information suggests ranges of several physical and chemical parameters that can be associated with at least three categories of stress: a) no biological stress caused by anthropogenically altered physical or chemical conditions, including nutrients (naturally varying populations); b) biological stress caused by anthropogenically elevated nutrients but not by physical or non-nutrient chemical conditions; and c) biological stress caused by a number of anthropogenically altered, competing physical and/or chemical conditions, which may include elevated nutrients. If the connections between specific levels of macroinvertebrates and periphyton status categories or "tiers" in Maryland low order streams, and possibly macroinvertebrates and phytoplankton in Coastal Plain streams and rivers, can be documented using monitoring program parameters, the periphyton and phytoplankton monitoring data could be incorporated into the stressor identification process as new indicators of nutrient degradation. This has the potential enhance the Maryland capability to detect nutrient impairment.

7.4 Recommendations for Additional Monitoring and Analysis

This section presents our recommendations for potential additional analyses that could be performed to supplement the analyses described in this report. Some of these analyses attempt to close the gap between the stressor-response analyses which relate nutrients to primary production and Maryland's water quality standards. Other analyses attempt to shed light on the relation between nutrients and the IBIs that are used to measure aquatic life support in 1st - 4th order streams. Finally, the successful identification of stressor-response relations suggest the potential for using those relations to characterize the trophic status of Maryland's waters in a nutrient biotic index and to develop BCG tiers along with criteria for the highest tiers.

7.4.1 Recommendations for Relating Primary Production by Phytoplankton or Periphyton to Maryland's Water Quality Standards

Sections 4 and 5 have identified relations between nutrients and chlorophyll *a* in phytoplankton and periphyton, respectively. As discussed in Section 7.2, Maryland does not have water quality standards for chlorophyll *a* for either phytoplankton or periphyton in free-flowing streams. Maryland does have chlorophyll *a* criteria for drinking water in reservoirs, but it has not been established that the rationale for those criteria apply to free-flowing streams.

The rationale for the drinking water chlorophyll *a* criteria is based on the determination that chlorophyll *a* concentrations above 30 µg/liter represent a shift in algal taxa to blue-green algae, and that an average growing season concentration of 10µg/liter can be correlated with excursions of the instantaneous concentration above 30 µg/liter. An analysis should be conducted in free-flowing streams to determine if a level of chlorophyll *a* concentration can be associated with a shift to blue-green algae or other nuisance taxa, and whether the frequency of nuisance taxa blooms can be related to an average chlorophyll *a* concentration in the longer term or to another statistical measure of long-term concentration. Two data sets exist that have both monitored chlorophyll *a* concentrations and information on algal species, cell counts, and other measures characterizing phytoplankton. The Washington Aqueduct has collected this information at their intakes on the Potomac River. The Ohio River Valley Water Sanitary Commission (ORSANCO) also has a long-standing algal and nutrient monitoring program that can furnish the type of information necessary to explore the relation between chlorophyll *a* concentrations and shifts in algal species in flowing waters.

For large rivers, where there is no established methodology in Maryland to determine aquatic life use support, identifying a threshold in chlorophyll *a* concentrations where the shift to nuisance species occurs may also be sufficient to provide a rationale for establishing nutrient criteria to protect aquatic life. Otherwise, development of a large river assessment methodology will be necessary before excess primary production in large rivers, as represented by chlorophyll *a* concentrations, can be related to the general health of the biological community. For smaller order rivers and streams, it is necessary to relate chlorophyll *a* concentrations associated with phytoplankton or periphyton with the biocritieria for aquatic life use. Currently in Maryland, there is insufficient data to relate measures of periphyton biomass, such as chlorophyll *a* content, to macroinvertebrate metrics, as was done for nutrients in this study. It may be possible, however, to relate benthic metrics (as the dependent variable) to monitored periphyton chlorophyll *a* or nutrient concentrations using the Virginia or SRBC periphyton datasets. We propose that such analyses be attempted using the methods of this study, in particular, using RPART to determine the impact of confounding variables. If such analyses are successful, recommendations can be made to establish a periphyton monitoring program in Maryland to collect the data necessary to support such analyses. It may also be possible to relate water-column chlorophyll *a* concentrations to benthic metrics in small-order streams in the Coastal Plain using existing data. It would be necessary to relate the location and timing of monitored chlorophyll *a* concentrations to biological sampling. We recommend that the feasibility of associating monitored chlorophyll a concentrations with biological monitoring in the Coastal Plain be explored and, if an association can be established, to proceed with an RPART analysis to determine the chlorophyll *a* breakpoints that impact biological metrics.

7.4.2 Recommendations for Clarifying the Relation Between Nutrients and MBSS IBI Scores

Although our analyses established that many benthic metrics are sensitive to nutrients, the MBSS BIBI, which is one of the two main biocriteria in Maryland, was not sensitive to nutrients (Sections 6.4.2 and 7.2.4). One of the possible reasons why the MBSS BIBI did not show a response to nutrients may be due to the differences between the reference conditions used to define the BIBI and the MDL samples set used in our analyses. If the list of reference sites and their attributes could be obtained from DNR, comparing those reference sites to the MDL samples and their characteristics would help shed light on the relative insensitivity of the BIBI to nutrients. It could help determine, for example, whether the MBSS reference conditions or the MDL conditions are more restrictive, or whether they have a more complicated reason to produce differing results.

The nutrient response of fish metrics in general and the MBSS FIBI in particular were not analyzed in this project. Fish monitoring data comparable to those of the MBSS program are difficult to find. Since MBSS benthic samples turned out to dominate the filtered dataset, it is possible that the response of the MBSS FIBI to nutrients in the filtered data set could also be explored in a future analysis.

7.4.3 Nutrient Biotic Index

Biological monitoring is used to measure the health of the biological community, but it can also be used to measure water quality conditions. Different benthic, fish, or algae taxa tend to have different optimal conditions with respect to water quality characteristics such as dissolved oxygen, pH, organic pollution, or nutrients. It follows that the relative abundance of different taxa are an indicator water quality conditions. The closer the pH is to a taxa's optimum pH, for example, the more likely that taxa can be found in relatively high abundance. One standard way to calculate an index is by weighing the taxa optima by their relative abundance. Nutrient indices have been constructed in this manner using diatoms (Van Dam *et al*., 1994) and benthic macroinvertebrates (Smith *et al*., 2007). Since the biological community is relatively stable, while measurements of water chemistry can vary on very short time scales, biological indices can be better indicators of water quality conditions than water quality monitoring.

This study has demonstrated that benthic metrics and chlorophyll a concentrations in phytoplankton and periphyton are sensitive across nutrient gradients that represent distinct ecological conditions. It may be possible to use these metrics and chlorophyll concentrations, rather than the nutrient optima of distinct taxa, to develop of nutrient index for Maryland's free-flowing waters.

7.4.4 Development of a stressor gradient scale

The vertical axis of the BCG diagram (**Figure 32**) and the biological characteristics associated with different tiers (**Table 28**) were originally designed with stream macroinvertebrates and fish in mind. Maryland monitoring programs have assessment criteria, categories of impairment, and a Tier II designation for high quality streams that contain much of the information needed to develop boundaries for at least some of the BSG tiers. The corresponding stressor scale on the x-axis of the BCG diagram, or the GSG, is more difficult to develop. Multiple, inter-related anthropogenic stressors affect biological communities and untangling their various competing impacts in order to identify or quantify stress attributable to one of the variables is challenging. If such a GSG scale were developed, however, it would facilitate efforts to link environmental conditions in the stream corridor to anthropogenic activities in the watershed. The GSG scale effectively becomes the y-axis of a new diagram relating land and water uses to stream conditions.

This study's nutrient bins, protective nutrient thresholds, and associated thresholds for "filtering" or accounting for different environmental parameters can provide some of the elements needed to create a multi-variable stressor gradient, or GSG. Such a gradient could be useful in Maryland's ongoing efforts to establish a tiered system of identifying and classifying streams according to stress. The effects of physiographic region (include karst geology) and Strahler stream order characterized in the study would be useful for developing the ecotypes for the individual BCG-GSG matrices. Guided by the recursive partitioning analysis results, the relative importance of different habitat and water quality stressors can be determined, scaled, and incorporated into the stressor gradient. Habitat and water quality levels as well as nutrient concentrations protective of high quality biological communities were recognized in the study. These conditions form the basis for the "upper end" of the GSG scale and would be associated with one of the higher BCG tiers.

7.4.5 Summary of Recommendations

To summarize, we recommend the following analyses be performed as part of the next steps in nutrient criteria development:

- Obtain Washington Aqueduct algal monitoring data and analyze relation between chlorophyll *a* concentrations and algal taxa;
- Obtain ORSANCO algal monitoring data and analyze relation between chlorophyll *a* concentrations and algal taxa;
- Using the methods of this study, analyze relation between benthic metrics as the dependent variable and periphyton chlorophyll *a*, phosphorus, or AFDM as independent variables using VADEQ and SRBC periphyton monitoring data;

- Try to associate chlorophyll *a* monitoring data in the Coastal Plain with macroinvertebrate monitoring data and, if successful, determine relation between macroinvertebrate metrics and observed chlorophyll *a* concentrations;
- Explore relation between MBSS reference conditions and MDL samples used in this study;
- Test sensitivity of MBSS FIBI to nutrient bins;
- Develop Nutrient Biotic Index using sensitive benthic metrics, chlorophyll a water column concentrations, and or chlorophyll a periphyton concentrations; and
- Explore potential to develop BCG tiers and a stressor gradient which incorporates the habitat, water quality, and biological thresholds identified in this study

8. Implications of Numerical Nutrient Criteria for the Chesapeake Bay TMDL Allocations

With regard to nutrient criteria for free-flowing rivers and streams, Maryland is perhaps currently in a unique position among the states. Nearly the entire state, with the exception of a portion of Garrett County that drains to the Ohio River and the portion of the Eastern Shore that drains directly to the Atlantic Ocean, is subject to nitrogen and phosphorus TMDLs for the Chesapeake Bay. Under the Chesapeake Bay TMDLs, sources of nitrogen and phosphorus will be subject to stringent controls. **Table 29** shows the reductions in nitrogen and phosphorus from current conditions required for full implementation of the Chesapeake Bay TMDLs. These reductions are on top of the 33% reduction in nitrogen loads and 38% reduction in phosphorus loads achieved statewide since 1985 under the voluntary strategies of the Chesapeake Bay Program and before the adoption of the Bay TMDLs in 2010 (MDE 2010c).

Needless to say, it will be a considerable challenge to meet the Bay TMDL implementation goals by 2020, as Maryland has pledged to do, and to maintain the TMDL loading caps in the face of continued population growth in the region. The costs are substantial. For example, Maryland's Watershed Implementation Plan (WIP) calls for 68 major Wastewater Treatment Plants (WWTP), processing 95% of the state's wastewater, to upgrade to Enhanced Nutrient Removal (ENR). These upgrades, as well as upgrades to Blue Plains, the major WWTP in the Washington metropolitan area, will require an over $4 billion investment (MDE 2010c). Phase I of the WIP also estimates that meeting the Bay TMDL allocation goals by 2020 will require almost $4 billion of stormwater system improvements by jurisdictions subject to municipal separate stormwater sewer system (MS4) permits. Rural areas of the state will also bear the expense of meeting the nutrient reduction requirements of the Bay TMDLs. Over 400 million of septic system improvements are part of Maryland's implementation plan. Adoption of additional agricultural best management practices (BMPs) and other changes in agricultural operations are expected to cost in the neighborhood of $200 million.

Table 29. Estimated nutrient reductions required to meet 2017 interim targets (70% of final) and 2020 final allocations for the Chesapeake Bay TMDL (% reduction from current loads as represented by CBP 2009 Progress Scenario), by Sector.

Section	% Nitrogen Reduction from 2009 Progress to meet 2017 interim targets	% Phosphorus Reduction from 2009 Progress to meet 2017 interim targets	% Nitrogen Reduction from 2009 Progress to meet 2020 final Bay TMDL allocations	% Phosphorus Reduction from 2009 Progress to meet 2020 final Bay TMDL allocations
Regulated Urban Stormwater	9%	12%	18%	34%
Non-regulated Urban Stormwater	-7%	-4%	19%	39%
Agriculture	6%	3%	23%	12%
CAFO	20%	28%	12%	31%
Septic	26%	0%	39%	0%
Forest	0%	0%	0%	0%
Air	-1%	-1%	1%	2%
WWTP & CSO	39%	34%	26%	21%
Total	**16%**	**12%**	**21%**	**18%**

Given the cost, schedule, and level of effort required to implement the Bay TMDLs, it is important to anticipate both the impact of the Bay TMDLs on local water quality and the potential impact of the adoption of nutrient criteria on Bay TMDL implementation. Are the Bay TMDL nitrogen and phosphorus allocations sufficient to protect not only the Bay but also local free-flowing rivers and streams? If they should prove insufficient, what level of effort would be required to protect aquatic life in local rivers and streams from harmful nutrients impacts? The planning horizon for major upgrades to WWTPs and stormwater systems is on the order of 10 to 20 years, and the state is unlikely to obtain the cooperation of regulated entities if major potential changes in the regulatory environment, like the adoption of numerical nutrient criteria, introduce an additional element of uncertainty into the implementation process.

Unfortunately, at the time this document is being written, it is not possible to resolve whether Bay TMDLs will be sufficient to protect local free-flowing streams from potential harmful impacts of nutrients, primarily because it is unknown at this time what implementation of the Bay TMDLs will look like on a local scale. WIPs are being developed in three phases. The Phase I Plan was released concurrently with the Bay TMDLs in 2010. It provides the information necessary for broad allocation goals to meet the nitrogen, phosphorus, and sediment TMDL for the 98 water quality limited segments that make up Chesapeake Bay. It also specifies an interim goal that in Maryland’s case is a commitment to achieve 70% of the final TMDL reductions by 2017. A Phase II Plan will be developed in 2012 that will specify nutrient reduction targets and implementation goals at a more local level, at the county scale. The Phase III Plan will be developed in 2017. It will specify the strategy for meeting the final TMDL reductions, which Maryland has pledged to meet in 2020, five years ahead of the deadline for full implementation of the Bay TMDLs.

8.1 Problems Testing Implications of Numerical Nutrient Criteria with Phase 5 Watershed Model

The Phase II WIPs will be based on a revised version of the Chesapeake Bay Program (CBP) Phase 5 Watershed Model. Phase 5.3.0 was used to set the Bay TMDLs in 2010. A revised version, Phase 5.3.2, had been calibrated at the time this report was written but no reduction scenarios associated with the Bay TMDL have been released.

The P5 model is a Hydrological Simulation Program FORTRAN (HSPF) model of Maryland, Virginia, and portions of Pennsylvania, New York, and West Virginia that are within the Chesapeake Bay basin. Its primary purposes are (1) to determine the sources of nitrogen, phosphorus, and sediment to the Chesapeake Bay, (2) to calculate nutrient and sediment loads to the Chesapeake Bay for use in the CBP model of water quality in the Bay, and (3) to estimate nutrient and sediment load allocations under nutrient and sediment TMDLs for impaired Chesapeake Bay segments. Bicknell *et al.* (2001) describe the HSPF model in greater detail. US EPA (2010b) documents the development of the P5 model.

While the P5 Model would provide the best available prediction of nitrogen and phosphorus concentrations under the Bay TMDL allocations, two features of the model hinder using it to investigate the relation between potential numerical nutrient criteria and the Bay TMDL. The first feature is the scale of the P5 Model. Generally, the model represents river reaches that have average annual flows greater than 100 cubic feet per second (cfs). MDE has worked with CBP to ensure that the main reach of all of MD’s non-tidal 8-digit watersheds are represented in the model, but generally speaking, the model does not represent rivers and streams smaller than 4th order.

Second, the calibration objectives for the P5 Model do not lend themselves to capturing concentrations of nutrients, particularly phosphorus, under ambient flow conditions. Broadly speaking, the P5 Model calibration has two objectives: (1) the distribution of simulated concentrations should match the distribution of observed concentrations; and (2) the P5 Model nutrient and sediment loads should match the loads calculated by the U. S. Geological Survey's River Input Monitoring Program (RIM) using USGS's statistical software, ESTIMATOR, which calculates loads based on regression equations that relate concentrations to flow, time, and season.

To meet these goals, not only are river simulation parameters adjusted but edge-of-stream nutrient loads are adjusted using multipliers called "regional factors." To meet the first objective, model parameters are adjusted to match the simulated and observed distribution curves by minimizing the error in the average of the log of the concentrations by quintile. **Figure 33** illustrates this procedure. **Table 30** gives the parameters adjusted based on quintile biases. In the case of total phosphorus, for example, the model is calibrated against the top three quintiles of observed data by adjusting the refractory matter settling rate and the phosphate concentration in scoured sediment, on the assumption that (a) most of the phosphorus load is transported during storm flow, and (b) there is a strong correlation between phosphorus concentrations and flow.

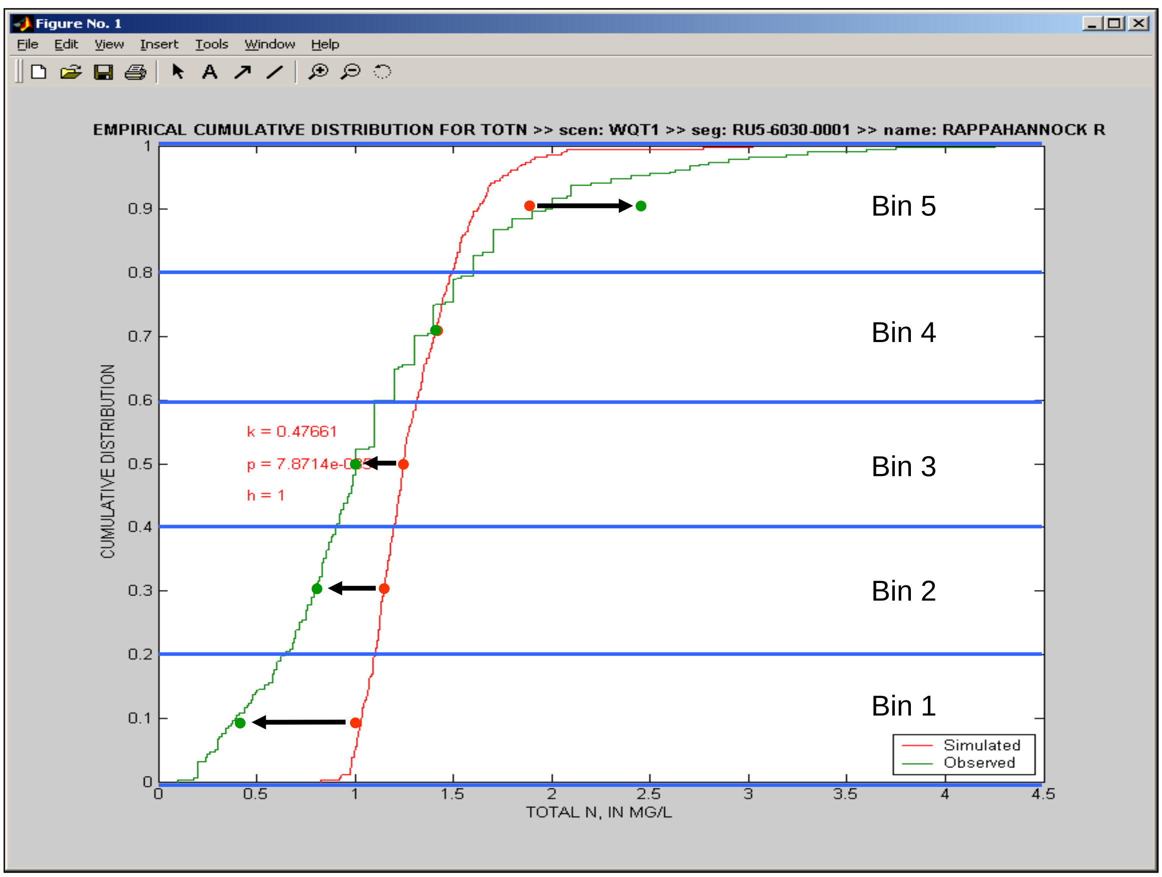


Figure 33. Cumulative Frequency Distribution Quintiles and Calculation of Biases. (Source: G. Shenk, CBP Modeling Subcommittee Meeting 10/17/2006).

Table 30. Phase 5 River Calibration Parameters and Cumulative Frequency Distribution Quintile Targets

Constituent	Parameter	Description	Quintiles
CHLA	PYSET	Phytoplankton settling rate	1-3
	MAGR	Maximum phytoplankton growth rate	4-5
TN, TP	REFSET	Organic matter settling rate	3-5
NO3	KNO320	Denitrification rate	1-3
NH4	KTAM20	Nitrification rate	1-5
TN	BPNH4	Sediment ammonia concentration (rivers only)	4-5
TP	BDPO4	Sediment phosphate concentration (rivers only)	4-5
PO4	ADSPO4	Adsorption coefficient for PO_4	4-5
TN	BRTAM	Ammonia release rate (lakes only)	1-5
TP	BRPO4	Phosphate release rate (lakes only)	1-4

If there is not a RIM station downstream of a reach, or another location where the USGS has calculated ESTIMATOR loads, regional factors are also applied to adjust input loads to reaches to reduce bias in the distribution of simulated concentrations. With the exception of the Choptank River, there are no ESTIMATOR loads calculated for the Eastern and Western Shore watersheds. If there is a RIM station or another calibration point with ESTIMATOR loads downstream of a reach, regional factors are used to adjust input loads to obtain better agreement between simulated nitrogen and phosphorus loads and their ESTIMATOR counterparts.

These calibration goals can conflict, as the frequently do in the Potomac River basin, where to match the ESTIMATOR loads calculated for the Fall Line monitoring station at Chain Bridge requires regional factors greater than one. This leads to an over-simulation of nutrient concentrations for much of the distribution curve. **Figure 34** illustrates the over-simulation of phosphorus in the lower quintiles of the distribution on the Monocacy River at Bridgeport.

8.2 Adapting CBP Assessment Methodology to Test Implications of Numerical Nutrient Criteria

The Chesapeake Bay Program has developed a methodology for taking into account model error when comparing modeling results with water quality criteria. CBP's Chesapeake Bay Water Quality and Sediment Transport Model (WQSTM) is used to simulate water quality constituents such as dissolved oxygen, nutrients, and chlorophyll *a* in the Chesapeake Bay. WQSTM is first calibrated against current and historical conditions to simulate observed concentrations of DO and other constituents. The calibrated model is then used to simulate management scenarios that represent different loading rates of TN, TP, and sediment, primarily to determine the water quality response to reduced nutrient and sediment loads. TMDLs for the Bay water quality segments are determined using these scenarios, but because the model does not accurately represent DO concentrations at all locations in the bay, model simulations are not compared to DO criteria directly. Instead, CBP uses a two-stage procedure with the nickname "scenario-izing the observations" (US EPA 2010b, Linker *et al.* 2002).

In the first stage of the procedure, linear regression models are determined at each sampling location where DO is monitored. The linear regression models relate simulated DO concentrations at that location, as the independent variable, to simulated concentrations at the same location under the load reduction scenario. In the second stage, the water quality response at a location under a management scenario is predicted by applying the regression equation to the observed concentrations to generate

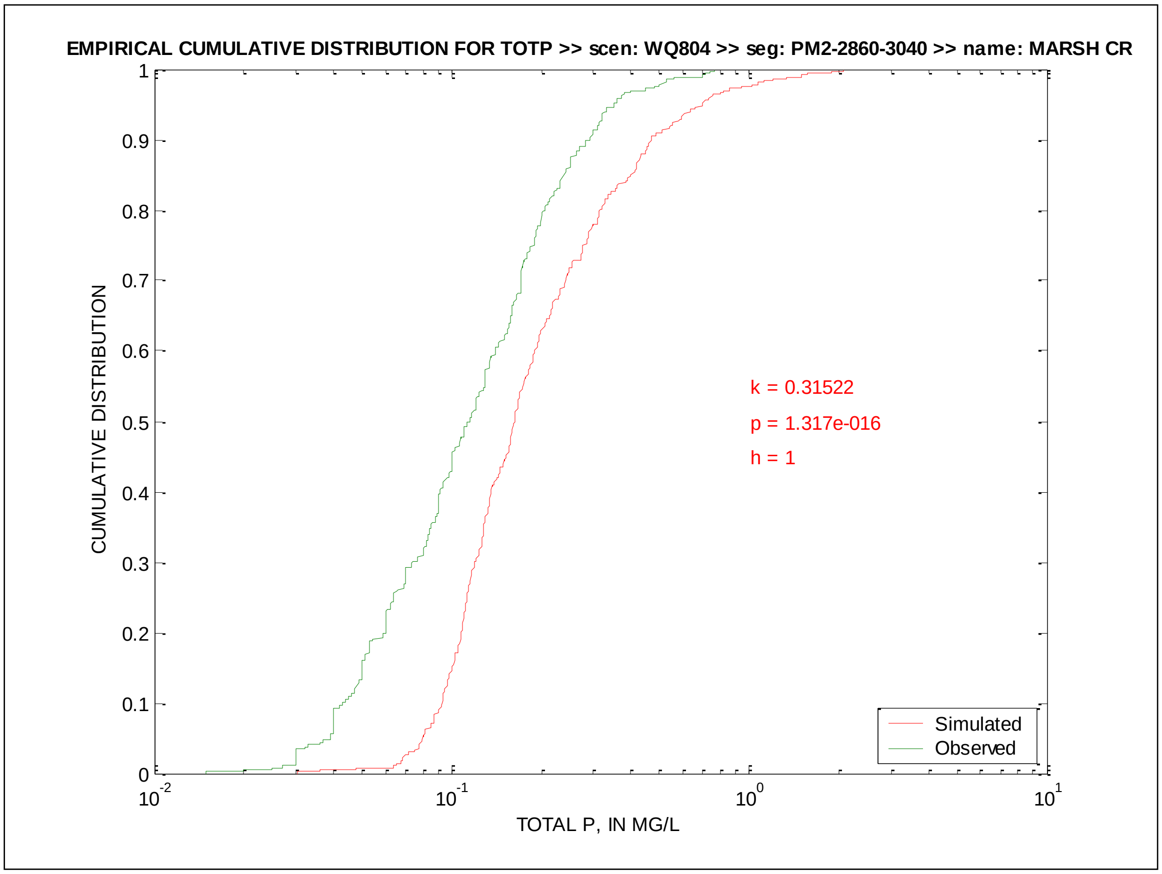


Figure 34. Observed and simulated total phosphorus, Monocacy River at Bridgeport, CBP Phase 5.2 Watershed Model.

"scenario modified" concentrations at that location. It is these scenario-modified concentrations that are compared to the DO criteria to determine if water quality standards are met. The transformed observations are past through another tool, the Interpolator, which, as the name suggests, interpolates in observed concentration space and time to assess standards in Chesapeake Bay water quality segments. When the Interpolator is used with observed concentrations, it assesses whether standards are met under current conditions. When it is used with transformed concentrations, it assess whether standards would be met under a management scenario.

A similar procedure can be applied to determine how the Bay TMDL nutrient reductions would affect nitrogen and phosphorus concentrations in free-flowing non-tidal waters, at least for 8-digit watersheds that include all or at least most of their headwaters. For those watersheds, it is plausible to assume that the simulated river reaches representing those segments represent average conditions in the watershed, and integrate water quality impacts on the watershed scale. Under these conditions, a two-stage procedure could be used to predict nutrient concentrations under the Bay TMDL allocations and to assess whether the biological health of local streams are protected from the impact of excess nutrients. The first stage would be to develop a regression equation relating simulated concentrations under the Phase 5 calibration to simulated concentrations under the Bay TMDL Allocation Scenario. The second stage would be to use the regression model to transform the observed concentrations in an 8-digit watershed and compare the transformed concentrations to nutrient thresholds.

8.3 Trial Analysis

The Maryland 8-digit Upper Pocomoke River watershed (02130203) was used to illustrate the procedure described above. The Upper Pocomoke River watershed lies in Wicomico and Worcester Counties on Maryland's Eastern Shore. **Figure 35** shows the location of the watershed. The watershed covers approximately 95,500 acres. Approximately 53% of watershed is forest or wetland, 39% agriculture, and 8% developed (MDE 2010a).

The Upper Pocomoke River was identified on Maryland's 2008 Integrated Report as impaired by nutrients, sediment, and impacts to biological communities (MDE 2008). Of the 23 sites monitored during Round 1 and Round 2 of the MBSS Program, eight had failing FIBI and/or BIBI scores. According

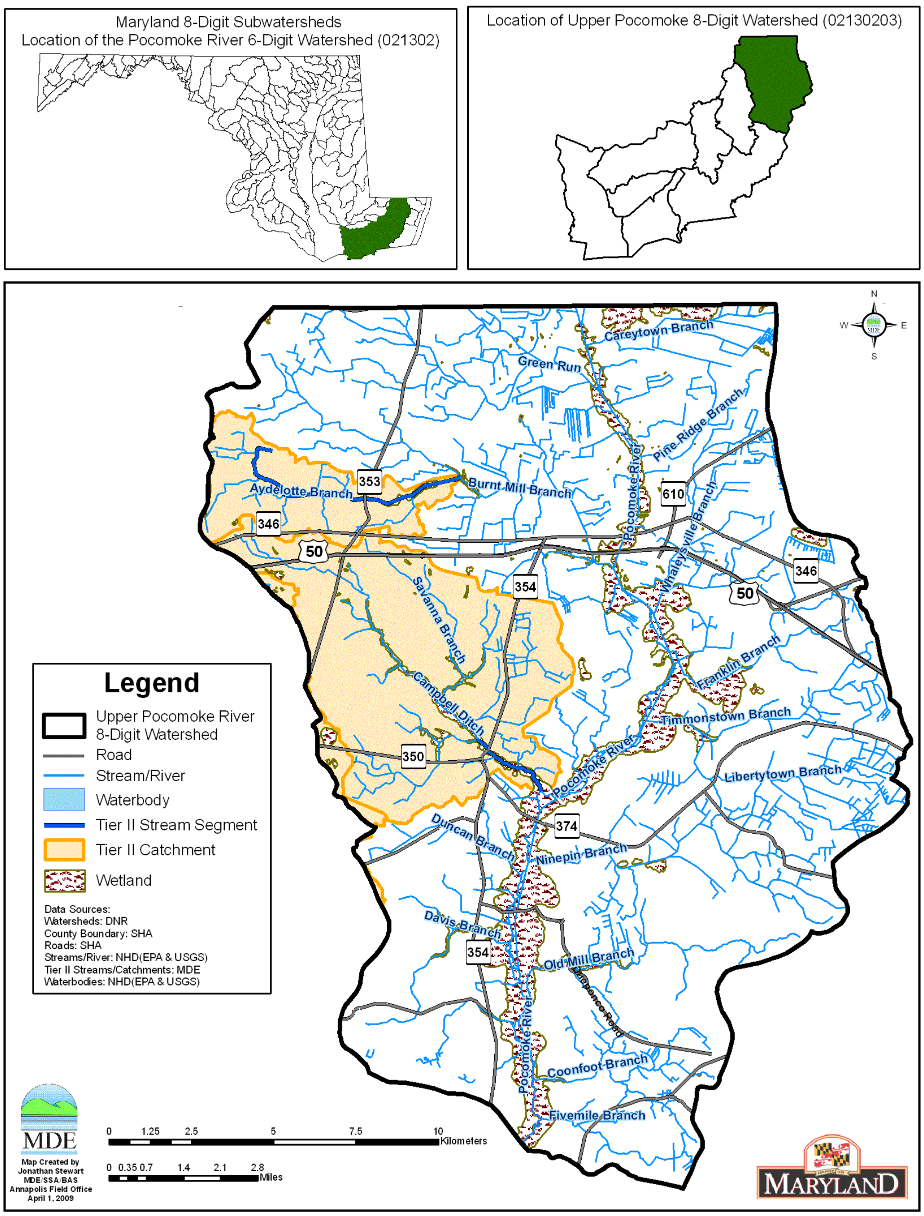


Figure 35. Location of the Upper Pocomoke River Watershed (MDE 2010a).

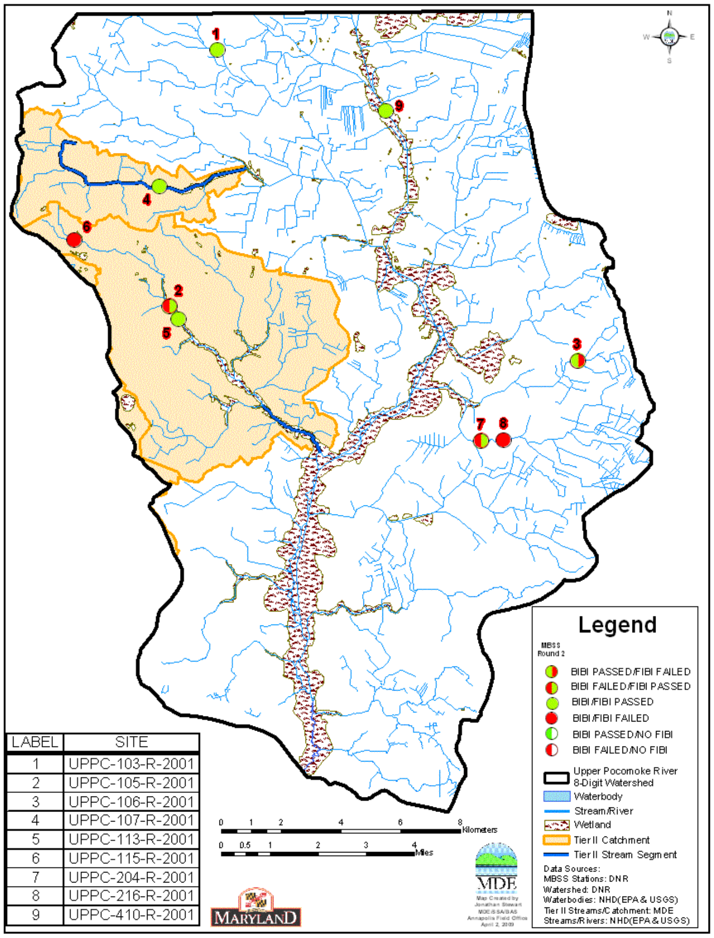


Figure 36. MBSS Round 2 Sampling Sites in the Upper Pocomoke River Watershed (MDE 2010a).

to the Biological Listing Methodology (BLM), it is estimated that 35% of the stream miles in the watershed are degraded.

A draft Biological Stressor Identification (BSID) report on the Upper Pocomoke River watershed (MDE 2010a) identified stressors and sources of the biological impairment. **Figure 36** shows the location of the Round 2 MBSS sampling sites used in the draft report. **Table 31** summarizes the biological monitoring results under MBSS Round 2. High total phosphorus and high orthophosphate were identified as stressors, impacting an estimated 37% and 68%, respectively, of the impaired stream miles in the watershed. Sediment, in-stream habitat, low dissolved oxygen, and high sulfates were also identified as stressors. High total nitrogen was not identified as a stressor. Identified sources were associated with agricultural activities, either in the riparian buffer or in the watershed in general.

The Upper Pocomoke watershed was chosen to illustrate the scenario-ization method only because the Upper Pocomoke River model segments were the only segments where daily simulated concentrations were available for a Phase 5.3.0 Model scenario representing the implementation of the nutrient reduction strategies in Maryland's Phase I WIP (Scenario 2010MDWIP10N11210). Only the method is

Table 31. Upper Pocomoke River watershed MBSS Round 2 biological monitoring results with TP concentrations (mg/liter) observed and predicted under Phase 1 WIP.

				Sample	Total Phosphorus (mg/liter)	
Site	FIBI	BIBI	Status	Date	Observed	Predicted WIP-1
UPPC-103-R-2001	3.50	2.71	FAIL	3/12/2001	0.0298	0.030
UPPC-105-R-2001	3.25	1.57	PASS	3/13/2001	0.0809	0.064
UPPC-106-R-2001	1.50	2.14	PASS	3/15/2001	0.0762	0.061
UPPC-107-R-2001	4.50	3.29	FAIL	3/6/2001	0.0396	0.037
UPPC-113-R-2001	3.00	2.43	FAIL	3/13/2001	0.0533	0.046
UPPC-115-R-2001	2.50	1.57	PASS	3/8/2001	0.1164	0.085
UPPC-204-R-2001	2.50	1.86	PASS	3/15/2001	0.1942	0.126
UPPC-216-R-2001	2.00	2.14	PASS	3/15/2001	0.1879	0.123
UPPC-410-R-2001	3.00	3.57	FAIL	3/13/2001	0.0577	0.049

represented here: no conclusions can or should be drawn from the results.

The Pocomoke River is represented by two river segments in the Phase 5 Watershed Model: EL2_5110_5270 and EL2_5270_0001. **Figure 37** shows the location of the model segments. Only segment EL2_5110_5270 was used to develop the regression relations, because it was calibrated against observed data, although in theory multiple segments could be used.

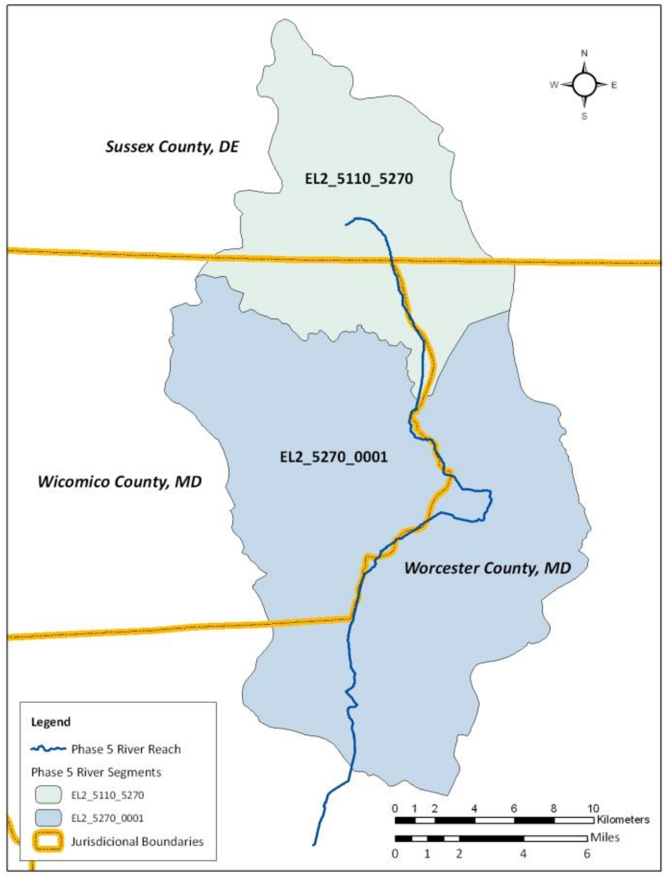


Figure 37. Chesapeake Bay Program Phase 5 Watershed Model Segments representing the Pocomoke River.

Orthophosphate concentrations were larger in the WIP scenario than in the Calibration Scenario, so no regression was performed for orthophosphate. It is worth repeating that no conclusions should be drawn from this. The fact that orthophosphate concentrations under the WIP Scenario are larger than under the calibration could be (1) an error, (2) a peculiarity of the Phase 5.3.0 Model, corrected in Phase 5.3.2, or (3) a peculiarity of the Pocomoke River simulation.

A regression equation was estimated for TP. The CBP methodology estimates a separate regression equation for each month of the simulation. Since, as shown in Table 1, all TP samples were collected in March, 2001, the regression equation was confined to that month. The simulated concentrations were log-transformed before estimating the regression relation, because phosphorus concentrations can be expected to be log-normally distributed. **Figure 38** plots the simulated TP concentrations for the calibration against the simulated concentrations under the WIP1 Scenario. The estimated regression equation is:

Log TPWIP1 = 0.7747*log TPCalibration – 0.3471

with a coefficient of determination (R2) of 0.79.

The observed TP concentrations collected in MBSS Round 2 and used in the BSID analysis were then transformed using the regression relation. This provides a prediction of the TP concentrations under the phosphorus reductions required under the Phase I WIP. Observed and predicted TP concentrations are shown in Table 31. In this case, all predicted concentrations are below the BSID high TP threshold concentration of 0.14 mg/liter, in which case the proposed method would call for predicting that there would be no significant association between high total phosphorus concentrations and biological impairments in the Upper Pocomoke River watershed.

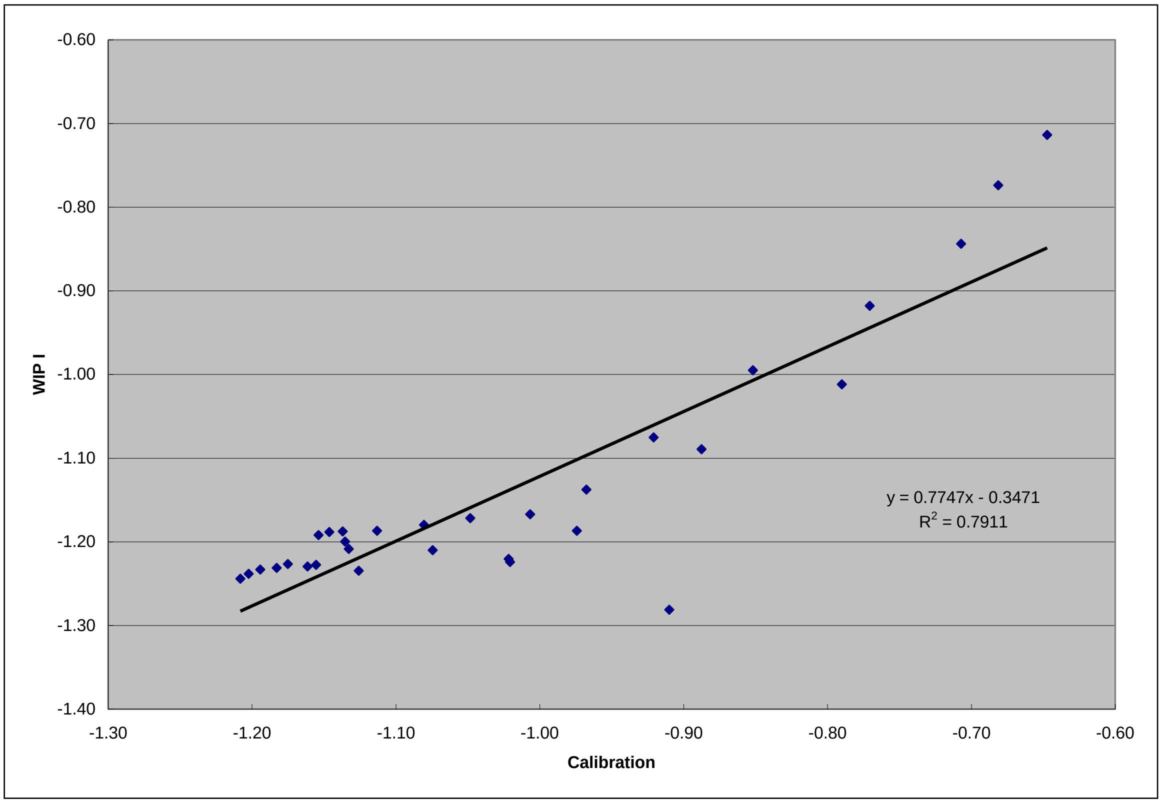


Figure 38. Simulated TP (mg/liter), March 2001, Phase 5.3.0 Calibration and 2010MDWIP1 Scenarios.

Again, this application is only meant to be an illustration of a method that could be used to predict what nutrient concentrations would be like under the full implementation of the Chesapeake Bay TMDLs, and to predict whether local water quality would be protected under the TMDL allocations. A test of the viability of the method would require (1) an examination of the performance of the revised Phase 5.3.2 Model under both the calibration and management scenarios, and (2) the application of the method to a variety of 8-digit watersheds, not just the Upper Pocomoke River. With so much at stake in the implementation of the Chesapeake Bay TMDLs, it may be worth exploring this and other avenues for predicting the impact of the required nutrient and sediment reductions on local water quality.

9. Literature Cited

Allen, J. D. 1995. Stream ecology: structure and function of running waters. Chapman & Hall, NY. 388p.

Baier, T., and E. Neuwirth. 2007. Excel :: COM :: R. Computational Statistics 22:91-108.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Bicknell, B.R., J.C. Imhoff, J.L. Kittle, Jr., and A.S. Donigian, Jr. 2001. Hydrological Simulation Program Fortran (HSPF): User's Manual for Release 12.

Buchanan, C., R. V. Lacouture, H. G. Marshall, M. Olsen, and J. M. Johnson. 2005. Phytoplankton Reference Communities for Chesapeake Bay and its Tidal Tributaries. Estuaries 28: 138-159.

Buchanan, C., K. Foreman, J. Johnson, and A. Griggs. 2011. Development of a Basin-Wide Benthic Index of Biotic Integrity for Non-Tidal Streams and Wadeable Rivers in the Chesapeake Bay Watershed: Final Report to the Chesapeake Bay Program Non-Tidal Water Quality Workgroup. ICPRB Report 11-1. Prepared for the US Environmental Protection Agency, Chesapeake Bay Program. Available: http://www.potomacriver.org/cms/publicationspdf/ICPRB11-01.pdf

Burton, J. and J. Gerritsen. 2003. A Stream Condition Index for Virginia in Non-coastal Streams. Prepared for USEPA Office of Science and Technology, USEPA Region III, and VA Department of Environmental Quality by Tetra Tech, Inc.

Canton, S. P., and J. W. Chadwick. 1983. Seasonal and Longitudinal Changes in Invertebrate Functional Groups in the Dolores River, Colorado. Freshwater Invertebrate Biology 2:41-47.

Carlson, R. E. 1977. A Trophic State Index for Lakes. Limnology and Oceanography 22:361-369.

Chen, P. 2011. Personal communication.

Cross, W. F., J. B. Wallace, A. D. Rosemond, and S. L. Eggert. 2006. Whole-System Nutrient Enrichment Increases Secondary Production in a Detritus-Based Ecosystem. Ecology 87:1556-1565.

Davies, S. P. and S. K. Jackson. 2006. The Biological Condition Gradient: A Descriptive Model for Interpreting Change in Aquatic Ecosystems. Ecological Applications 16:1251-1266.

Dzialowski, A. R., D. H. Huggins, F. J. deNoyelles, V. H. Smith, N. C. Lim, B. A. Baker, and J. Beury. 2009. Development of Predictive Models for Geosmin-Related Taste and Odor in Midwestern U. S. Drinking Water Reservoirs. Water Research 43:2829-2840.

Fisher, T. R., and A. B. Gustafson. 2003. Nutrient-Addition Bioassays in Chesapeake Bay to Assess Resources Limiting Algal Growth. Progress report: August 1990–December 2002. Prepared for Maryland Department of Natural Resources, Chesapeake Bay Water Quality Monitoring Program, by University of Maryland Horn Point Laboratory, Cambridge, MD.

Foreman, K., C. Buchanan, and A. Nagel. 2008. Development of Ecosystem Health Indexes for Non-tidal Wadeable Streams and Rivers in the Chesapeake Bay Basin. Report to the Chesapeake Bay Program Non-Tidal Water Quality Workgroup. December 5, 2008

Gerritsen, J. and E. W. Leppo. 2005. Biological Condition Gradient for Tiered Aquatic Life Use in New Jersey. Tetra Tech, Inc. Owings Mills, MD.

Gordon, N. D., T. A. McMahon, and B. L. Finlayson. 1992. Stream Hydrology: an Introduction for Ecologists. Wiley & Sons, NY.

Grumbles, B. 2007. May 25, 2007 Memo on Nutrient Pollution and Numeric Water Quality Standards. USEPA, Washington DC. http://water.epa.gov/scitech/swguidance/standards/criteria/nutrients/memo2007.cfm

Hawkins, C. P., and J. R. Sedell. 1981. Longitudinal and Seasonal Changes in Functional Organization of Macroinvertebrate Communities in Four Oregon Streams. Ecology 62:387-397.

Karr, J. R. and D. R. Dudley. 1981. Ecological Perspective on Water Quality Goals. Environmental Management 5:55-68.

Kirk, J. T. O. 1994. Light and Photosynthesis in Aquatic Ecosystems. Cambridge University Press, NY.

Lampert, W. and U. Sommer. 1997. Limnoecology: The Ecology of Lakes and Streams. Translated by J. F. Haney. Oxford University Press, NY.

Linker, L. C., G. W. Shenk, P. Wang, C. F. Cerco, A. J. Butt, P. J. Tango, and R. W. Savidge. 2002. A Comparison of Chesapeake Bay Estuary Model Calibration with 1985-1994 Observed Data and Method of Application to Water Quality Criteria. USEPA Chesapeake Bay Program Office, Annapolis, MD.

Lorenzen, M. W. 1980. The Use of Chlorophyll-Secchi Disk Relationships. Limnology and Oceanography 25:371-372.

Mandel, R., A. Griggs, A. Rettig, and T. Fox. 2010. Nutrient Criteria Literature Review. Draft report prepared by the Interstate Commission on the Potomac River Basin for Maryland Department of the Environment, January 2010.

Maryland Department of the Environment (MDE). 2008. The 2008 Integrated Report of Surface Water Quality in Maryland. Maryland Department of the Environment, Baltimore, MD. Available online at: http://www.mde.state.md.us/programs/Water/TMDL/Integrated303dReports/. (Accessed March 2011).

Maryland Department of the Environment (MDE). 2009. Maryland Biological Stressor Identification Process. Maryland Department of the Environment, Baltimore, MD.

Maryland Department of the Environment (MDE). 2010a. DRAFT Watershed Report for Biological Impairment of the Upper Pocomoke River Watershed, Wicomico and Worcester Counties, Maryland: Biological Stressor Identification Analysis Results and Interpretation. Maryland Department of the Environment, Baltimore, MD.

Maryland Department of the Environment (MDE). 2010b. Rationale for Proposed Chlorophyll a Criteria for Maryland Drinking Water Reservoirs. Maryland Department of the Environment, Baltimore, MD.

Maryland Department of the Environment (MDE). 2010c. Draft. Maryland's Phase I Watershed Implementation Plan for the Chesapeake Bay Total Maximum Daily Load. Maryland Department of the Environment, Baltimore, MD:

Megard, R. O., J. C. Settles, H. A. Boyer, and W. S. Combs, Jr. 1980. Light, Secchi-Disks, and Trophic States. Limnology and Oceanography 25:373-377.

Merritt, R. W., and K. W. Cummins. 1996. An Introduction to the Aquatic Insects of North America (3rd edition). Kendall/Hunt Publishing Co. Dubuque, IA.

Morgan, R. P., II, and K. M. Kline. 2010. Nutrient concentrations in Maryland non-tidal streams. Environ. Monit. Assess. DOI 10.1007/s10661-010-1684-0.

Olivero, A. P. and M. G. Anderson. 2008. Northeast Aquatic Habitat Classification System. The Nature Conservancy. Report prepared in conjunction with the Northeast Association of Fish and Wildlife Agencies. Available: http://www.natureserve.org/publications/pubs/NE_Hab_Class&Map_0708_FinalRept.pdf

Omernik, J. M. 1995. Ecoregions: A Spatial Framework for Environmental Management, in Davis, W.S. and Simon, T.P. (eds.), Biological Assessment and Criteria, Tools for Water Resource Planning and decision making. Lewis Publishers, Boca Raton, FL, pp. 31-47.

Ormerod, S. J., and R. W. Edwards. 1987. The Ordination and Classification of Macroinvertebrate Assemblages in the Catchment of the River Wye in Relation to Environmental Factors. Freshwater Biology 17:533-546.

Paul, M. J., and L. Zheng. 2007. Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania: TMDL Application. Report prepared by Tetra Tech for the U.S. Environmental Protection Agency, Region III, November 2007.

Perrin, C. J., and J. S. Richardson. 1997. N and P Limitation of Benthos Abundance in the Nechako River, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 54:2574-2583.

Plafkin, J. L, M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish. EPA/440/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Available: http://www.epa.gov/owow/monitoring/rbp.

Robertson, D., D. J. Graczyk, P. J. Garrison, L. Wang, G. LaLiberte, and R. Bannerman. 2006. Nutrient Concentrations and their Relations to the Biotic Integrity of Nonwadeable Rivers in Wisconsin. Professional Paper, U.S. Geological Survey, Reston, VA.

Robertson, D. M., B. Weigel, and D. Graczyk. 2008. Nutrient Concentrations and their Relations to the Biotic Integrity of Nonwadeable Rivers in Wisconsin. Professional Paper, U.S. Geological Survey, Reston, VA.

Roth, N. E., M. T. Southerland, J. C. Chaillou, P. F. Kazyak, and S. A. Stranko. 2000. Refinement and Validation of a Fish Index of Biotic Integrity for Maryland Streams. CBWP-MANTA-EA-00-2. Maryland Department of Natural Resources, Annapolis, MD. Available online at: http://www.dnr.state.md.us/streams/pdfs/ea-00-2_fibi.pdf

Scientific Advisory Board (SAB). 2010. Scientific Advisory Board Review of Empirical Approaches for Nutrient Criteria Derivation. EPA-SAB-10-006. U.S. Environmental Protection Agency, Washington, DC.

Sheeder, S. A., and B. M. Evans. 2004. Estimating Nutrient and Sediment Threshold Criteria for Biological Impairment in Pennsylvania Watersheds. Journal American Water Resources Association 40:881-888.

Smith, A. J., R. W. Bode, and G. S. Kleppel. 2007. A Nutrient Biotic Index (NBI) for Use with Benthic Macroinvertebrate Communities. Ecological Indicators 7:371-386.

Southerland, M. T., G. M. Rogers, M. J. Kline, R. P. Morgan, D. M. Boward, P. F. Kazyak, R. J. Klauda, and S. J. Stranko. 2005. New Biological Indicators to Better Assess the Conditions of Maryland Streams. CBWP-MANTA-EA-05-13. Maryland Department of Natural Resources, Annapolis, MD. Available online at: http://www.dnr.state.md.us/streams/pdfs/ea-05-13_new_ibi.pdf.

Southerland, M., J. Volstad, E. Weber, R. Morgan, L. Currey, J. Holt, C. Poukish, and M. Rowe. 2007. Using MBSS Data to Identify Stressor for Streams that Fail Biocriteria in Maryland. Maryland Department of the Environment, Baltimore, MD.

Stevenson, R. J., and L. L. Bahls. 1999. Chapter 6. Periphyton Protocols. In Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, Second Edition. US Environmental Protection Agency.

Stribling, J. B., Jessup, B. K., White, J. S., Boward, D., & Hurd, M. (1998). Development of a Benthic Index of Biotic Integrity for Maryland Streams. Maryland Department of Natural Resources, Annapolis, MD. CBWA-MANTA-EA-98-03. Available online at: http://www.dnr.state.md.us/streams/pdfs/ea-98-3_benthic_ibi.pdf

Stowe, W.2011. Personal communication.

Summers, J. 2008. Assessment of Filamentous Algae in the Greenbrier River and other West Virginia Streams. Report prepared for West Virginia Department of Environmental Protection. Available: http://www.dep.wv.gov/WWE/watershed/wqmonitoring/Documents/Greenbrier/Assessment_Filamentous_Algae_Greenbrie_%20River.pdf

Thorp, J. H., and A. P. Covich [eds]. 2001. Ecology and Classification of North American Freshwater Invertebrates (2nd edition). Academic Press.

USEPA. 1998. National Strategy for the Development of Regional Nutrient Criteria. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 2005. Use of Biological Information to Better Define Designated Aquatic Life Uses in State and Tribal Water Quality Standards: Tiered Aquatic Life Uses. DRAFT. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 2009. Empirical Approaches for Nutrient Criteria Derivation. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 2010a. Chesapeake Bay Total Maximum Daily Load. Appendix H: Criteria Assessment Procedures using Model Scenario Output with Bay Monitoring Data. Chesapeake Bay Program Office. Annapolis, MD.

USEPA. 2010b. In preparation. Chesapeake Bay Phase 5.3 Community Watershed Model. EPA XXX-XXX-10. U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis MD. December 2010. Available: http://ches.communitymodeling.org/models/CBPhase5/documentation.php#p5modeldoc

USEPA. 2010c. Using Stressor-Response Relationships to Derive Numeric Nutrient Criteria. EPA-820-S-10-001. U.S. Environmental Protection Agency. Washington, DC.

USEPA. 2011. Wadeable Streams Macroinvertebrate Methods Comparison [Microsoft Excel Book]. Revised April 5, 2011. USEPA Region III, Wheeling, WV. Contact: Kelly Krock, 304-234-0242, krock.kelly@epa.gov.

Van Dam, H., A. Mertens, and J. Sinkeldam. 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. Aquatic Ecology 28:117-133.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Vollenweider, R. A., W. Rast, and J. Kerekes. 1980. The Phosphorus Loading Concept and Great Lakes Eutrophication. In R. C. Loehr, C. S. Martin, and W. Rast [eds]. Phosphorus Management Strategies for Lakes. Ann Arbor Science Publs., Ann Arbor, MI.

Walker, W.W., Jr. 1984. Statistical Bases for Mean Chlorophyll a Criteria. Lake and Reservoir Management: Proceedings of Fourth Annual Conference. North American Lake Management Society, pp. 57-62.

Ward, J. V. 1992. Aquatic Insect Ecology. 1. Biology and Habitat. John Wiley & Sons, Inc.

Weigel, B., and D. Robertson. 2007. Identifying Biotic Integrity and Water Chemistry Relations in Nonwadeable Rivers of Wisconsin: Toward the Development of Nutrient Criteria. Environmental Management 40:691-708.

Wetzel, R. G. 2001. Limnology: Lake and River Ecosystems. Academic Press.

Wood, A. J., J. M. Olmernik, and D. D. Brown. 1999. Level III and IV Ecoregions of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. U.S. Environmental Protection Agency technical document, Corvallis, OR. 61pp. Available: http://www.epa.gov/wed/pages/ecoregions/reg3_eco.htm

Zheng, L., J. Gerritsen, J. Beckman, J. Ludwig, and S. Wilkes. 2008. Land Use, Geology, Enrichment, and Stream Biota in the Eastern Ridge and Valley Ecoregion: Implications for Nutrient Criteria Development. Journal of the American Water Resources Association 44:1521-1536.