HALL & ASSOCIATES

Suite 203
1101 15th Street, N.W.
Washington, D.C. 20005-5004
Web Site: http://www.hall-associates.com
Reply to E-mail:

Telephone: 202-463-1166

Fax: 202-463-4207

jhall@hall-associates.com

September 3, 2009

Via Email and U.S. Mail

Dr. Thomas Armitage
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Ave, NW
Mail Code 1400F
Washington, DC 20460

Re: Comments on EPA Guidance for Developing Numeric Water Quality Criteria for Nutrients

Dear Dr. Armitage,

Hall & Associates represents a group of Pennsylvania dischargers, known as the Pennsylvania Periphyton Coalition, that are impacted by EPA's proposed guidance entitled "*Empirical Approaches For Nutrient Criteria Derivation.*" As discussed in the enclosed comments, we believe the approaches suggested for nutrient criteria development are seriously flawed, will both over and under regulate nutrients, will fail to restore use impairments of concern and will waste state and local resources on a nationwide basis. Consequently, we respectfully request that the Science Advisory Board reject the new approaches to numeric nutrient criteria development recommended by the draft guidance.

Sincerely,

John C. Hall

Enclosures

cc: Pennsylvania Periphyton Group
John W. Brosious, Pennsylvania Municipal Authorities Association
Pennsylvania Water Environment Association

CRITICAL EVALUATION OF EPA'S DRAFT EMPIRICAL APPROACHES FOR NUTRIENT CRITERIA DERIVATION

By

John C. Hall, Esq. (Hall & Associates)
William T. Hall (Hall & Associates)
Dominic Di Toro, PhD, PE (University of Delaware)
Thomas Gallagher, PE (HydroQual, Inc.)

INTRODUCTION

The EPA Science Advisory Board (SAB) will meet on September 9 – 11, 2009 to consider the charge questions posed by the EPA Office of Water, Office of Science and Technology to review EPA's draft guidance on *Empirical Approaches For Nutrient Criteria Derivation* (hereafter "EPA Criteria Derivation Report"). The basic purpose of Section 303(c) criteria is to establish the level of water quality that is required to protect designated uses from specific adverse impacts associated with a pollutant. This document was released for public review on August 17, 2009 and it presents six statistical methods for analyzing nutrient data to relate the effect of nutrients to indicators of invertebrate use impairment. The guidance also presents a chapter on evaluating candidate stressor-response criteria from a consideration of the various statistical methods (See Guidance, Chapter 5 at 46). Together, these statistical approaches and the evaluation method are intended to serve as the primary basis for deriving Clean Water Act Section 303(c) numeric nutrient criteria. Where uncertainty exists in the regression analyses, EPA recommends that the lower confidence bound of the regression analysis be used to select the numeric nutrient criteria applicable to all similarly classified waters. EPA further indicates that even if an impairment threshold cannot be identified for a water body type, statistical "change point analyses" are sufficient to identify the necessary level of water quality that must be achieved. These criteria will serve as the basis for identifying waters as nutrient impaired and for preparing TMDLs to restore designated uses to nutrient-impaired waters.

The methods discussed in the draft EPA Criteria Derivation Report have already been used by EPA, as part of a "weight of evidence" analysis to derive numeric nutrient endpoints for several TMDL applications. (Attachment 1 (Piedmont End Point Report), Attachment 2 (Allegheny Plateau and Ridge and Valley Ecoregions of Pennsylvania: TMDL Application)). This "weight of evidence" analysis also considers distributional statistics and "literature" recommendations, in addition to stressor-response data, as the basis for selecting numeric nutrient endpoints for TMDL application. See EPA Criteria Derivation Report at 18. Objections raised on these TMDLs initiated an appeal of these methods and a request, supported by the Pennsylvania delegation, that the entire procedure undergo SAB review. (Attachment 3 (Request for SAB review) and Attachment 4 (support letter from Senators Specter/Casey)). In response, EPA agreed to conduct a SAB review of the new statistical methods being employed to derive nutrient criteria. EPA confirmed that SAB approval would allow nationwide implementation of the new approach (Attachment 5). In advance of this review, a detailed history of the

misapplication of these various statistical concepts in nutrient TMDL development was prepared by Hall & Associates and published in BNA. (Attachment 6). This analysis, which demonstrated that specific application of simplified methods misdirected resources and failed to identify appropriate solutions for invertebrate impairments, provides insight to the Board on why EPA is seeking SAB approval of these new methodologies.

It is axiomatic that the Section 303(c) water quality criteria be based on clearly demonstrated cause and effect relationships. "*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*", USEPA 1985. The draft EPA Criteria Derivation Report and the charge to the SAB seeks SAB confirmation that the stressor-response data and framework provide this demonstration. We believe that the statistical methods presented by EPA for evaluating the stressor-response data are wholly inadequate for deriving water quality criteria, will misdirect state and local resources on a nationwide basis and more comprehensive methods in a mechanistic framework are necessary to properly address nutrient impairment issues. Moreover, statistical methods do not demonstrate that a scientifically defensible cause-and-effect relationship exists between stressors and recognized use impairment metrics (e.g., distributional statistics). The analyses fail to address a host of well understood plant growth mechanisms as well as the other ecological factors that influence invertebrate population dynamics. What is presented is merely a correlation analysis and correlations do not prove causation. Moreover, the alleged correlations presented are demonstrably flawed. Therefore, these methods cannot serve as the basis for deriving either necessary or protective numeric water quality criteria.

CRITERIA DEVELOPMENT REQUIREMENTS

By statute, criteria must be based on the latest available science and set at the level necessary to protect aquatic life and human health uses (CWA Section 304(a)). To achieve this requirement it is essential that criteria possess two attributes: (1) the criteria must be based on data that confirm the pollutant is causing use impairment at ambient concentrations, and (2) the level at which the numeric criteria is set is both <u>sufficient</u> and <u>necessary</u> to protect designated uses. See, 40 CFR 131.2 (Purpose) and 131.3 (Definitions – "Criteria", "Section 304(a) Criteria"). Thus, criteria are, in general, set at the <u>threshold</u> level where the pollutant exposure is demonstrated not to pose a significant threat to aquatic life (Section 304(a); 40 CFR 131.2 131.3 (b), (c)).

Since 1985, EPA has had a well-defined procedure for developing scientifically defensible water quality criteria when it published the "*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*", USEPA 1985 (hereafter "*Guidelines*"). The *Guidelines* establish a number of very specific scientific screening procedures that must be met to establish criteria that meet Section 304(a) mandates, as follows:

- Water quality criteria must ensure use protection "with a small probability of considerable overprotection or under-protection." (*Guidelines* p. 5).

- It is not enough that the criterion is the best estimate given the available data. Criteria should be derived "only if adequate appropriate data are available to provide reasonable confidence that it is a good estimate." (*Guidelines* p. 5).

- Criteria must be based upon studies showing a clear close/response relationship to determine effect concentration. Data from confounded studies (i.e., results that are influenced by factors other than the pollutant of concern) should not be used. (*Guidelines* p. 15, 16, 21).

- All decisions should be based on a thorough knowledge of aquatic toxicology and criteria decisions must be altered if there is a substantial probability of over or under protection of aquatic organisms and their uses. (*Guidelines* p. 18).

- Based on "all available laboratory and field information", it must be determined that proposed criteria are "consistent with sound scientific evidence." If not, another criterion should be derived. (The concluding recommendation of the *Guidelines* p. 57).

The new approach recommended for nutrient criteria development, however, fails to reflect these long established principles. While the focus of the *Guidelines* requirements was on parameters that cause direct toxicity, the scientific principles summarized above apply to all criteria development, including to nutrient criteria. A demonstrated cause-effect relationship is a requirement of all criteria. The EPA Criteria Derivation Report nowhere makes such a demonstration with regard to invertebrate impacts. The regressions and statistical methods simply <u>assume</u> the nutrient plotted is the cause of the changed condition, even under circumstances where the nutrient levels could not rationally be expected to be causing increased plant growth. Regardless of the method of derivation, nutrient criteria must ensure use protection with a small probability of considerable overprotection or under-protection. If there is a substantial probability that the criteria derived using the proposed method are over or under-protective, such criteria must be rejected. It is apparent, however, that large uncertainties are associated with the suggested methods, as the nutrient concentration alleged to be causing the metric response may vary by a factor of 50. (See EPA Report Figures 13, 14, 16, 25). If the available data are insufficient to ensure that the resultant criteria are reasonable, those data need to be augmented or discarded in favor of more appropriate data. Finally, if field data confirm the approach is misplaced, it should not be utilized. Unfortunately, the recommended approach makes no allowance for consideration of actual conditions and will result in regulating nutrients even where it is apparent that the metric level is already achieved. (See, Figure 33).

Expert Opinion Surveys are Inadequate to Derive Criteria

The draft EPA Criteria Derivation Report addresses the development of numeric criteria in Section 3.1. In the first case, it notes that states may already have designated use criteria (e.g., biological criteria) that can be related to numeric nutrient levels. If such relationships can be developed that ensure the designated use criteria are achieved with a small probability of considerable overprotection or under-protection, the resulting numeric nutrient criteria would be appropriate. The draft EPA Criteria Derivation Report also provides an alternative case:

> Also, expert opinion regarding protective levels of variables can be methodically collected (Reckhow et al. 2005), and surveys can identify conditions that support user expectations for different waterbodies (Heiskary and Walker 1988). (EPA Criteria Derivation Report at 18).

Such an approach is not supported by Federal or State law and must be rejected as inappropriate for criteria development. Expert opinion surveys cannot replace specific requirements in federal and state law for deriving appropriate water quality criteria. The draft EPA Criteria Derivation Report is full of examples that derive candidate numeric nutrient criteria from metrics that are not recognized as a priori use impairments, particularly with regard to streams. These metrics include total taxa richness, number of EPT taxa, chlorophyll-a concentration, and diatom trophic state index. Such response metrics cannot be used as the basis for establishing numeric nutrient criteria unless those metrics are first designated as use impairment thresholds (i.e., criteria).

Failure to Consider Relevant Factors Influencing Plant Growth and Invertebrate Metrics

The EPA Criteria Derivation Report recognizes that biological metrics, in particular, may change dramatically in streams due to a host of conditions, natural and man-induced. (Report @ 14,17, 24, 25, 30, 34, etc.) Therefore, selection of a single metric is not scientifically defensible. The metric must be related to specific physical conditions that support the metric and its threshold. When criteria are met, water quality will generally protect the designated use. (See the definition of "criteria" in the draft EPA Criteria Derivation Report, at 76). If the methods contained in the draft EPA Criteria Derivation Report are appropriate, then compliance with the candidate numeric nutrient criterion will ensure that the biological metric is achieved. If the metric is not a use-impairment threshold (i.e., a criterion), then use restoration cannot be demonstrated and the *Guidelines* requirement is not met. However, it is not apparent that any of the examples presented in the draft EPA Criteria Derivation Report demonstrate an appropriate linkage between nutrients and the response variables. First, there is no linkage presented between plant growth and invertebrate impairment. Without this linkage, all other relationships are simply speculative.

In several cases presented, the stressor is identified as a nutrient concentration derived from a grab sample measurement (e.g., a measure of the instantaneous concentration at one point in time). (See, Figure 14). The response variable (e.g., total taxa richness) represents an instream condition that develops over an extended period of time unrelated to the time frame associated with the grab sample. In the case of stream macroinvertebrates, the community develops over time in response to antecedent conditions and the measure of total taxa richness reflects an averaging period consistent with the life cycle of the individual organisms, with some organisms living a year or more. Unless the grab sample measurement reflects the average condition over a period of time relevant to the biological response metric and other factors are <u>not</u> influencing the response metric (e.g., habitat), there is no reason to believe that the stressor measurement bears any relationship to the response. Since no evidence is presented in the EPA Criteria Derivation Report to suggest that stressor grab sample concentrations are representative of average conditions relevant to the response metric, the evaluation methods presented in the

draft EPA Criteria Derivation Report are not consistent with sound scientific evidence and cannot be used to derive criteria.

Finally, the draft EPA Criteria Derivation Report advocates the development of numeric nutrient criteria for nitrogen and phosphorus regardless of need. (See, Figure 33). The EPA Criteria Derivation Report ignores plant growth mechanics and suggests that reducing nutrient levels that are far above saturation growth levels will somehow limit plant growth and thereby improve the invertebrate metrics. This is not scientifically defensible. Moreover, while nitrogen and phosphorus are both needed to sustain growth if either one is sufficiently limited growth will be limited. This scientifically validated concept has served as the basis for nutrient control for decades and establishing criteria for both parameters where control on only one parameter is needed to restore designated uses guarantees that such an approach provides significant overprotection in violation of the *Guidelines*.

PROBLEM DEFINITION

The draft EPA Criteria Derivation Report provides a reasonable problem definition with regard to nutrients and water quality criteria. It notes:

> Nutrients are essential for plant and microbial growth and at natural concentrations are generally considered beneficial. Over-enrichment by nitrogen and phosphorus stimulates excessive rates of plant and microbial growth and can produce biological and physical responses in surface water that adversely affect water quality and aquatic life. (EPA, 2009 at 1).

Simply stated, nutrients are not toxic to aquatic life at the concentrations typically encountered in receiving waters. However, depending upon the physical setting, they can stimulate excessive plant growth which, under certain conditions, can cause designated use impairments with regard to aquatic life, drinking water supply, or recreation. Use impairment is <u>not</u> caused by an "excessive" nutrient concentration. Rather, use impairment (e.g., recreation, fishery) is attributed to the presence of excessive levels of plant growth. In addition, elevated levels of plant growth may cause aquatic life impairments (e.g., low dissolved oxygen, pH out of range, habitat impairment). However, the linkage between nutrients and the impairments associated with them is highly complex and affected by a multitude of factors as suggested in Figure 10 from the draft EPA Criteria Derivation Report (at 16; presented below).

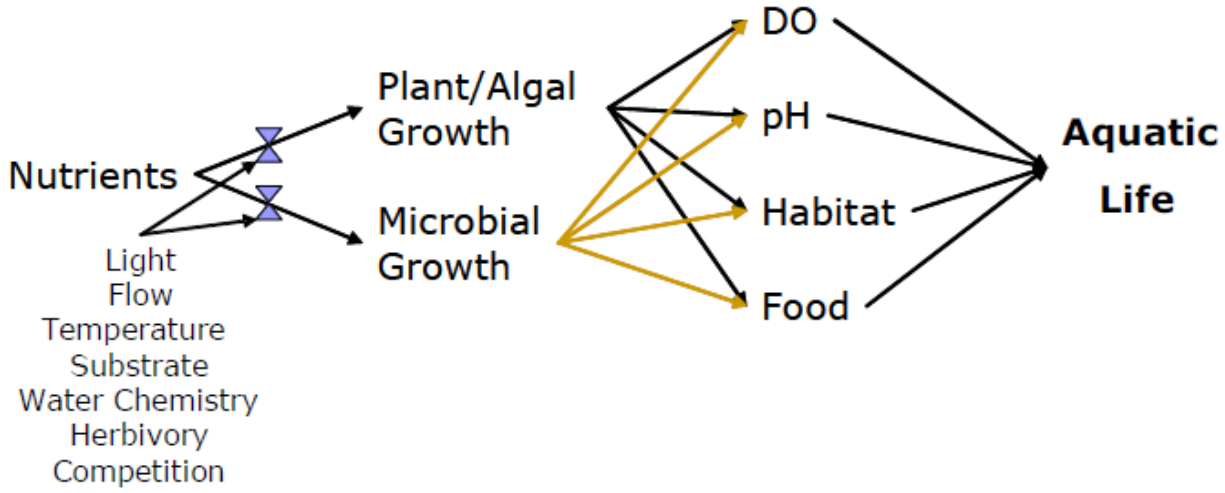


Figure 10. Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts. Nutrients enrich both plant/algal as well as microbial assemblages, which lead to changes in the physical/chemical habitat and food quality of streams. These effects directly impact the insect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

Thus, two steps are required to link nutrient to invertebrate impairments: (l) demonstrate that nutrients are causing the excessive plant growth, and (2) demonstrate the level of excessive plant growth that will cause invertebrate population impairment. The degree of understanding of this complex linkage depends upon the type of receiving water being considered and the site-specific habitat. In the case of lakes and other receiving waters with extended residence times, this understanding is more advanced (relatively speaking with respect to plant growth) and numerous mechanistic models exist that reasonably predict the level of phytoplankton growth in response to nutrient loading. The degree of variability (a key *Guidelines* criteria development issue) associated with phytoplankton chlorophyll-a levels and nutrients is about a factor of 4, as illustrated in Figure 12 from the draft EPA Criteria Derivation Report. This figure shows that the target chlorophyll-a level of 15 µg/L spans TP concentrations ranging from approximately 16 – 64 g/L . Mechanistic considerations of lake processes are required on a site-specific basis to reduce this variability and set appropriate TP targets for use restoration, because a given TP level (e.g., 35 ug/L) could easily produce both unimpaired (< 15u/L chl "a") or impaired (>15 ug/L chl "a") results. Thus, even for lakes, simply knowing the nutrient concentration present is not sufficient information to determine what, if anything, needs to be done though the range of uncertainty is relatively small.

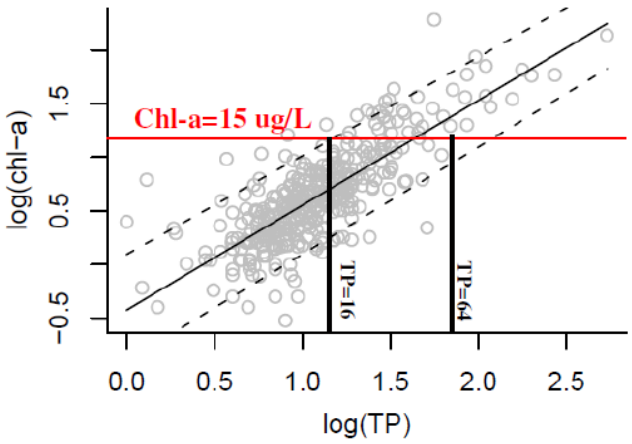
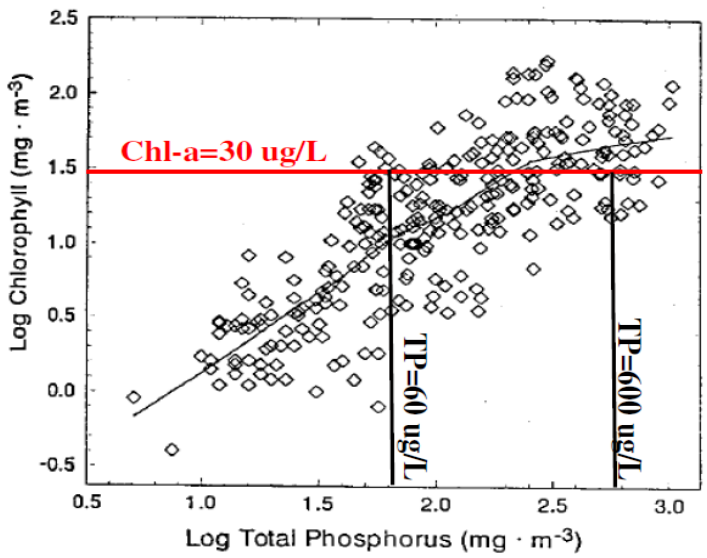


Figure 15. Log(TP) vs. log(chl *a*) for EMAP Northeast Lakes Survey. Dashed lines are the 5th and 95th percentile estimated by quantile regression. Solid line is the 50th percentile. Red line indicates where chlorophyll *a* = 15 µg/L. Units are µg/L.

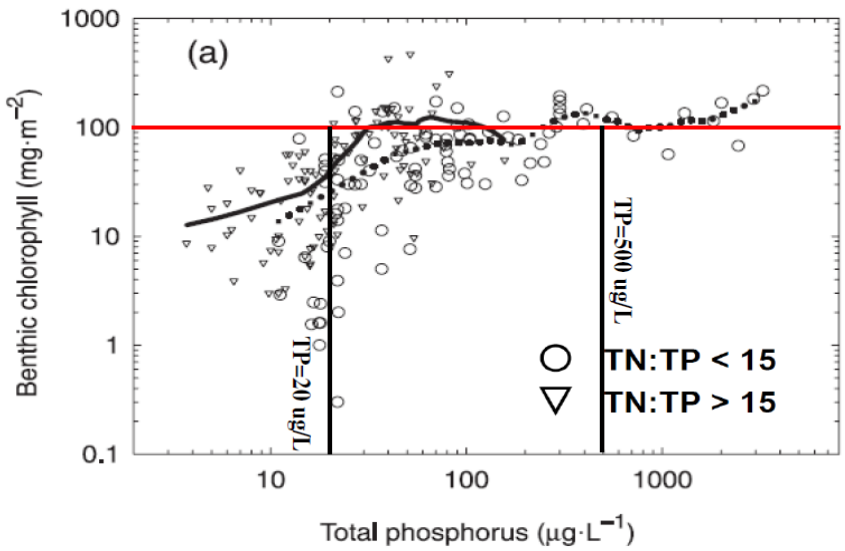
In the case of rivers and streams, the ability to predict plant growth responses is significantly reduced and the uncertainty in nutrient effects increases markedly. For large rivers with phytoplankton growth issues, compared to lakes, increases to about a factor of 10.



Total Phosphorus Vs. Benthic Chlorophyll-a
(From Dodds. W. K., Smith, V. H., and Lohman, K., 2006)

The work of Dodds regarding periphyton growth regressions was referenced in several places in the document, though the actual data and results were not presented. Dodds' work demonstrates that the uncertainty in periphyton growth for streams is far greater than phytoplantkton growth in rivers. The graph below indicates that periphyton responses vary by a factor of 25 with regard to nutrient level. While not presented in this report, EPA has extensively assessed the relationship between periphyton growth and nutrient levels. The results

generally confirm there is no demonstrable "regression" relationship between nutrient levels and periphyton response.



Total Phosphorus Vs. Benthic Chlorophyll-a
(From Dodds. W. K., Smith, V. H., and Lohman, K., 2006)

In fact, it was analyses such as these that led EPA to abandon its earlier recommendations that Dodd's simple regression equations be used to predict periphyton growth. Mechanistic models to predict the level of periphyton growth (i.e., benthic algae) are available and many factors that are unimportant in lakes and large rivers significantly affect such growth in streams (e.g., canopy, scour, substrate) are addressed in those model frameworks. This information confirms that there is no reasonable confidence in setting a river or stream nutrient objective, given the lack of cause and effect relationship presented by the periphyton data and wide range in phytoplankton responses. Thus, the following is apparent with regard to plant growth responses to elevated nutrient levels:

(1) simplified approaches used for lakes do not apply to rivers and streams, as the variability in response increases greatly for these waters; and,

(2) a scientifically defensible approach to rivers and streams must account for the factors influencing plant growth dynamics.

Regarding the "relationship" between nutrient levels and invertebrate populations, the uncertainty increases even further. This is evidenced by the extremely poor r^2 that results from the attempted correlation analyses (See, Figures 14, 16). In most instances, EPA does not even provide this basic statistical information for the SAB review. The poor correlation coefficient is indicative of a very weak relationship. For example, Figure 13 from the draft EPA Criteria Derivation Report illustrates total taxa richness in West Xeric region streams as a function of nutrient concentration. This figure shows that the target richness level of 40 spans TN concentrations ranging over two to three orders of magnitude (based on the 90% prediction

interval). Even at quite high nutrient levels (TN> 2 mg/l) taxa richness response ranges from less than 10 (very poor) to 70 (excellent). Such information confirms that some other factors, unrelated to nutrient level are responsible for these widely varying results. Similar results are found for phosphorus and EPT richness. See EPA Report Figure 14. This type of data scatter prevent the reasonable selection of a nutrient target level that is necessary to ensure use protection.

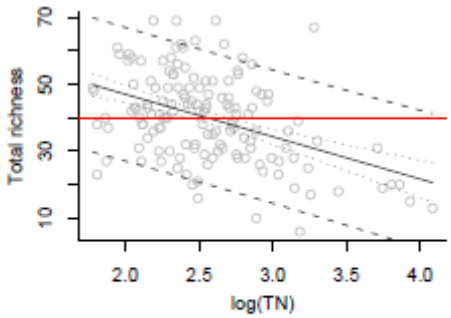


Figure 13. Log(TN) versus total species richness in EMAP-West Xeric region streams. Solid line: mean regression relationship, dashed lines: 90% prediction intervals. Red horizontal line indicates total richness = 40. Units are µg/L for log(TN). Regression equation: Total Richness = 72 – 13[Log(TN)],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

EPA'S REPORT IS MISSING CRITICAL FOUNDATION ANALYSES

The draft EPA Criteria Derivation Report does not provide any analysis showing the most critical relationship needed to establish an invertebrate-based nutrient criteria– the relationships between algal growth and invertebrate metrics. Moreover, the EPA Report is devoid of information showing a reliable connection between nutrient level and benthic plant growth for streams or rivers. The Report does cite on several occasions to the work of Dodds (relating periphyton growth to nutrient concentrations), though as shown above, such relationships are simply misplaced. (See also Attachment 6 confirming Dodds' equations fail to reliably predict periphyton growth in the PA TMDL action). This is rather dramatic oversight considering that nutrient criteria development is primarily targeted at excessive plant growth. If such a relationship was presented, it would likely show variability exceeding two orders of magnitude. This extreme level of variability is due to the fact that other factors tend to control plant growth in streams and invertebrate responses are more influenced by factors other than nutrients.

Tetra Tech (2008) (EPA's contractor on the EPA Criteria Derivation Report) conducted a literature review of nutrient – algal growth relationships for EPA in conjunction with the development of five nutrient TMDLs in Pennsylvania, and concluded the following:

> "Study results summarized as part of this literature review support the assertion that while a relationship may exist between periphyton growth and nutrients, the dynamics change as a function of multiple factors. These factors include antecedent conditions, water temperature, pH, light availability, flow regime, and grazing, among others. *Nutrient levels may be secondary to other determinants of biomass and growth such as light, disturbance, and grazing."* (Tetra Tech (2008a) at 18).

The linkage between nutrients and aquatic life is more indirect than that for plant growth. As depicted in Figure 10, algal growth and microbial growth are more directly affected by nutrients. These effects filter through the receiving water ecology and influence macroinvertebrate and fish communities. However, this linkage is more subtle because of the intervening steps that separate the animal communities from nutrients and the numerous other factors influencing the presence of these higher order organisms. For this reason, prior EPA nutrient criteria development guidelines have recommended that nutrient criteria development focus on the relationship between nutrients and plant growth (EPA, 2000). Once that relationship is defined, the more indirect impacts on macroinvertebrates and fish can be explored if necessary.

> "fish and macroinvertebrates do not directly respond to nutrients, and therefore may not be as sensitive to changes in nutrient concentrations as algal assemblages. 6It is recommended that relations between biotic integrity of algal assemblages and nutrients be defined and then related to biotic integrity of macroinvertebrate and fish assemblages in a stepwise, mechanistic fashion." *Nutrient Criteria Technical Guidance Manual – Rivers and Streams, USEPA July 2000* @ 85.

Contrary to the *Technical Guidance Manual*, the draft EPA Criteria Derivation Report provides multiple examples attempting to relate macroinvertebrate metric response directly to nutrient concentrations even though the response is far removed from the stressor and none of the necessary intervening plant growth responses were documented. Clearly, the candidate numeric nutrient criteria that result from this approach will be either too restrictive or not restrictive enough, and, with EPA's recommended approach, there will be no way of knowing which error has been made until the numeric criteria are achieved and the response is re-evaluated. Thus, it is apparent that the new recommended approach is guaranteed to misallocate resources on a large scale if applied to the regulatory process. Such an approach fails to meet *National Guidelines* prerequisites.

ESTABLISHMENT OF FIXED NUTRIENT CONCENTRATIONS AS CRITERIA IS INAPPROPRIATE

The expressed purpose of the draft EPA Criteria Derivation Report is to facilitate the development of nutrient criteria. Criteria are defined as constituent concentrations, levels, or narrative statements representing a quality of water that supports a particular use. When criteria are met, water quality will generally protect the designated use. With respect to Figure 10, the quality of water that supports a particular use is defined by the algal growth in the receiving water, constituent concentrations that may impair uses, affected by algal growth (i.e., DO, pH, clarity), and the aquatic life present. These measures are a direct indication of whether a designated use is impaired or not. The nutrient level is not the parameter that describes use impairment.

This situation is analogous to the relationship between dissolved oxygen and biochemical oxygen demand. Low dissolved oxygen level directly causes toxicity to aquatic life and water quality criteria are established to protect aquatic life by setting numeric DO requirements. Although BOD affects DO, there are no BOD water quality criteria because BOD, itself, is not

toxic and multiple factors affect how BOD is expressed in the water column. These factors are reasonably well known and site-specific, mechanistic models are used to account for them in setting appropriate effluent limitations. This well known approach should serve as a model for addressing nutrient-related impairments.

Consequently, nutrient criteria should be expressed as a level of algal growth or an aquatic life metric that reflects the designated use, impacted by algal growth (e.g., secchi depth in lakes, periphyton growth level that impairs invertebrate populations). Then, based on a mechanistic understanding of the waterbody, the concentrations of nitrogen and/or phosphorus necessary to achieve the target water quality may be ascertained because there are no uniform system responses to this pollutant. Mechanistic models that relate the significant determinants of criteria response must be applied to ensure the correct cause of impairment and solution are related. Then, the controls necessary to eliminate use impairment, as evidenced by the numeric criteria, can be implemented on a site-specific basis.

CONCLUSIONS AND RECOMMENDATIONS

The statistical procedures presented in the EPA Report are not a scientifically defensible substitute for developing a clear, reproducible relationship between a pollutant and the stressor of concern. Such information must exist to develop scientifically defensible Section 303(c) criteria. Moreover, it is apparent that extreme uncertainty exists under EPA's suggested approach such that there is no reasonable basis to believe that nutrients are either (1) the cause of a macroinvertebrate impairment or (2) that reducing nutrients to the levels described in the report would remedy such impairment. Where regression analyses are completed, the r^2 values all appear to be 0.1 or less. Thus, more than 90% of the "impairment" is NOT explained by the nutrient level. Using alternative statistical tools does not "improve" the lack of relationship contained in the dataset. The scientific community generally recognizes that such weak relationships are not a basis for concluding that a cause and effect relationship has been established. Such weak relationships cannot be used to develop necessary, sufficient and scientifically defensible water quality criteria.

Perhaps the greatest flaw in the suggested approach is that it assumes that nutrients are the cause of the impacts being assessed, ignoring (1) other factors that influence the metric being measured (e.g., habitat and other common stressors) and (2) known plant growth kinetics that confirm the measured impacts at high nutrient levels cannot be the cause of reduced invertebrate populations. In particular, where the saturation growth rate concentration is exceeded (generally above 50 ug/l TP or 300 ug/l TN), increased nutrient concentrations are not expected to cause additional plant growth. For these high nutrient levels, the sometimes measured dramatic decrease in invertebrate metrics must be caused by a co-occurring phenomena – such as excessive sedimentation or habitat alteration, not nutrients. Ignoring these well established factors that may control the type and richness of invertebrate populations is not scientifically defensible. Conversely, the analysis entirely failed to assess how the invertebrate metrics change in the range where nutrients may be the primary factor controlling plant growth – assuming excessive plant growth is occurring. Such analyses, as discussed herein, indicate that there is no demonstrable nutrient: invertebrate population relationship in this range of nutrient concentration.

Lastly, it is apparent from the extensive stream and river databases that attempts to select a single nutrient value to be applicable in all such waters is not rational. Plant growth responses, the primary concern of nutrient control, are not uniform, and are primarily controlled by the physical setting. Even in lakes where light is not generally limited and sufficient detention time is not an issue, selection of a single target value will allow the actual plant growth level to vary by a factor of 4. Selecting the target nutrient level from the lower end of the uncertainty range, as suggested by EPA, will ensure that a very substantial number of situations are over regulated. For streams, the range of plant growth responses increases to a factor of 20 or greater. Invertebrate responses that are even more remotely related to nutrient levels have a *2 order of magnitude range*. A more rational approach would be to identify appropriate plant and biological response levels that represent impaired and unimpaired conditions over a range of habitat types. Such indices could be used to appropriately identify waters that need remediation. Based upon the factors influencing the impairment, as appropriately assessed with site-specific information, the proper remedial measures would be identified. Where "excessive" plant growth is occurring, particularly with streams, the solution may not be nutrient reduction. Rather, various forms of canopy or stream bank restoration or flow/depth modification may be the more effective and possibly the only solution to address the condition. Again, this would be determined on a case-specific basis, not through a one-size-fits-all nutrient criteria that is certain to misallocate and misidentify impairments and appropriate solutions.

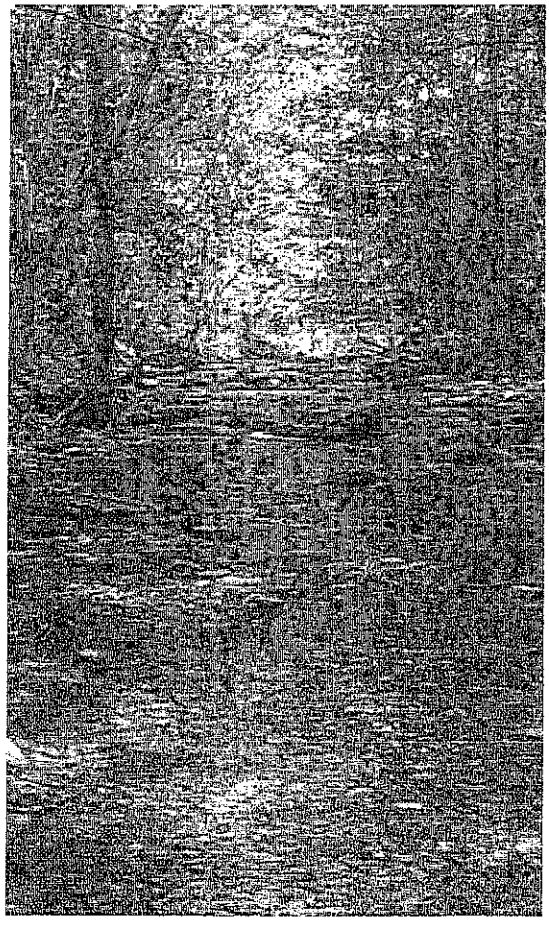
Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania: TMDL Application

Prepared for

United States Environmental Protection Agency

Region 3

Philadelphia, PA



By

Michael J. Paul and Lei Zheng

Tetra Tech, Inc.

400 Red Brook Boulevard, Suite 200

Owings Mills, MD 21117

November 20, 2007

Table of Contents

Page

Introduction 1

Frequency Distribution Based Approach 4

Modeled Reference Expectation Approach 7

N:P Ratios Suggest P Limitation Dominates the Northern Piedmont Ecoregion 7

Stressor-Response Approach 8

Data 10

Data Analysis: Overview 11

Data Analysis: Metric Calculation 13

Results: Algal Biomass – Nutrient Relationships 15

Results: Algal Metrics – Nutrient Relationships 16

Results: Benthic Macroinvertebrate Metrics – Nutrient Relationships 18

Literature Based Analysis: Current Existing Endpoints or Threshold Values 22

Recommended Endpoints 26

Total Phosphorus (TP) 26

Total Nitrogen (TN) 30

Literature Cited 34

Appendix A. Spearman Correlation matrices among enrichment related Environmental variables and a few macroinvertebrate metrics in Northern Piedmont Ecoregion in EMAP Database 37

Appendix B. Spearman Correlation coefficients among enrichment related Environmental variables and a sub-sample of macroinvertebrate metrics from Northern Piedmont Ecoregion stream samples in the USGS NAWQA database 38

List of Figures

Figure Page

1 Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts 1

2 Map of the sample sites used in the development of nutrient endpoints using the distribution based approach in this study, labeled by agency affiliation 5

3 Plot of total phosphorus samples in the All Sites and two Reference Site populations used to estimate candidate with the distribution based approach 6

4 Site average Chl a concentrations in relation to average TN and TP concentrations in the Northern Piedmont ecoregion 16

5 Response of diatom trophic state index (MSU_TSI) to total phosphorus concentrations In depositional habitats and riffles 19

6 Response of EPT taxa metric to the phosphorus gradient in the Northern Piedmont Ecoregion 20

7 Response of % clingers in benthic samples along phosphorus gradient in the Northern Piedmont Ecoregion 21

8 Response of % intolerant urban taxa in benthic samples along phosphorus gradient In the Northern Piedmont Ecoregion 22

List of Tables

Table Page

1 Values of TN and TP candidate endpoints derived using the distribution based approach 6

2 Values for molar N:P ratios estimated from the All Sites and the two Reference Site populations developed for use in this study 8

3 Biological data and their related chemical measurements in the stations in Northern Piedmont ecoregion 11

4 Spearman Correlation matrices between environmental parameters and algal metrics 17

5 Threshold values for the MBSS benthic macroinvertebrates IBI metrics in the Northern Piedmont ecoregion 20

6 Comparison of New England Water Quality Recommendations, after Mitchell et al. (2003) with EPA recommended regional endpoints 25

7 Summary of candidate endpoints for each of the analytical approaches discussed 27

Introduction

The United States Environmental Protection Agency (USEPA) in Region 3 is overseeing the development of nutrient Total Maximum Daily Loads (TMDLs) to protect aquatic life uses for several streams in the Northern Piedmont ecoregion of southeastern Pennsylvania. Specifically TMDLs are being developed for the following watersheds: Chester, Indian, Neshaminy, Skippack, Southampton, and Wissahickon Creeks. Tetra Tech, Inc (Tt) was approached to assist USEPA in establishing appropriate TMDL endpoints for nutrients that are both protective of aquatic life uses in this region and defensible. This document describes the process that was applied, the results of those analyses, and recommended nutrient endpoints for the TMDLs in question.

Nutrients affect aquatic systems in diverse ways, and the effects on most non-primary producer aquatic life uses are indirect (Figure 1).

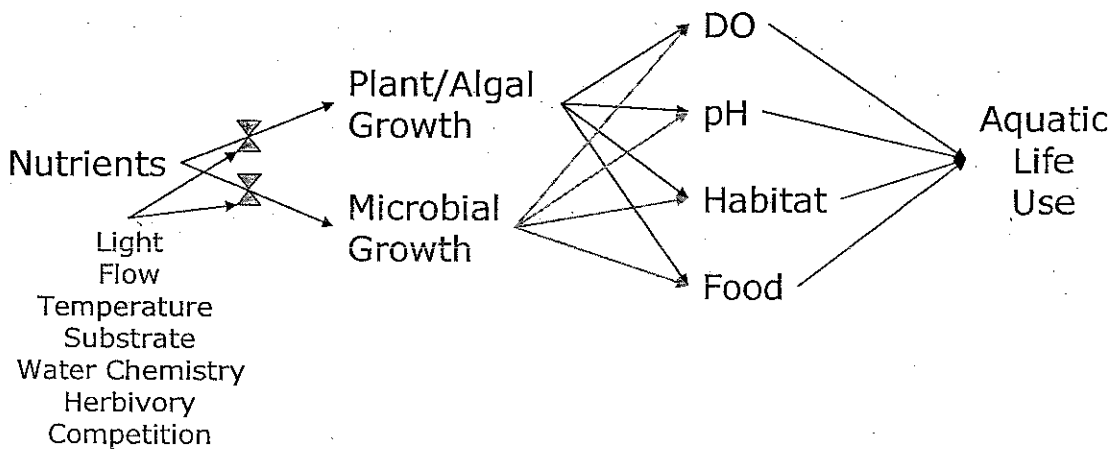


Figure 1 – Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts. Nutrients enrich both plant/algal as well as microbial assemblages, which lead to changes in the physical/chemical habitat and food quality of streams. These effects directly impact the insect and fish assemblages. The effects of nutrients are influenced by a number of other factors as well, such as light, flow, and temperature.

Nutrients cause enrichment of primary producer and decomposer biomass and productivity, the increase of which leads to changes in the physical and chemical stream environment (e.g., reduced oxygen, loss of reproductive habitat, alteration on the availability of palatable algal taxa, etc.). It is these effects which directly result in changes to the biological stream community (e.g., loss of disturbance sensitive taxa), and ultimately impair the use of a stream for aquatic life.

Traditionally, water quality endpoints to protect aquatic life use were developed using toxicological approaches. Such approaches have been applied for a range of pollutants to develop water quality endpoints, for example. However, as explained above, nutrient enrichment does not have a direct toxicological effect on non-primary producer aquatic life. It is worth mentioning that nutrients do, however, affect algal and plant aquatic life directly, altering the diversity and composition of those assemblages radically. For insects, fish and other aquatic life, however, the mode of action of nutrients is indirect and through a causal pathway that involves alteration of physical, chemical, and biological attributes of their habitat. As a result, traditional toxicological approaches are not appropriate.

The USEPA has published guidance on nutrient endpoint development for the protection of designated uses for a range of waterbody types including rivers and streams (USEPA 2000a), but also for lakes and reservoirs (USEPA 2000b), estuaries (USEPA 2001), and wetlands (USEPA 2007). The principal method described in those documents is the use of a frequency distribution-based approach (often called the reference approach), where a percentile of a distribution of values is used to identify a nutrient endpoint. The sample distributions were typically either from least disturbed reference

sites (sensu Stoddard et al. 2006) or the entire population of sample sites. These documents, however, clearly encourages the use of alternative scientifically defensible approaches and, especially, the application of several approaches in a multiple-lines-of-evidence framework, to establish defensible and protective endpoints. The document states that, "a weight of evidence approach that combines (multiple) approaches...will produce endpoints of greater scientific validity." The approaches recommended include the frequency distribution approach, stressor-response analyses, and literature based values.

In determining nutrient endpoints for developing TMDLs to protect aquatic life uses of northern piedmont streams in southeastern Pennsylvania, we relied on a multiple lines of evidence approach using all of the following approaches: frequency distribution based analysis, stressor-responses analyses, and literature based values. The following sections describe these approaches in detail including the methods used for each and the results. The resulting candidate values were then considered and a weight-of-evidence selection process applied to develop final endpoint recommendations.

Due to the limitation of watershed sizes and the difficulty in obtaining stressor response gradients (especially for reference sites) in the six target watersheds, we proposed using an ecoregional nutrient endpoint development approach similar to that applied for nutrient criteria development to identify nutrient targets that would protect aquatic life uses in these watersheds. The USEPA, in their recommendations for nutrient endpoint development, specified that "Ecoregional nutrient criteria will be developed to account for the natural variation existing within various parts of the country." (USEPA 2000a)

They go on to explain the importance of ecoregions:

"Ecoregions serve as a framework for evaluating and managing natural resources. The ecoregional classification system developed by Omernik (1987) is based on multiple geographic characteristics (e.g., soils, climate, vegetation, geology, land use) that are believed to cause or reflect the differences in the mosaic of ecosystems."

The six targeted watersheds are located within the Northern Piedmont ecoregion. We collected data from across the same ecoregion but used data from only selected sites within Pennsylvania, Maryland, and New Jersey—three states that have similar geology to the six watersheds. We also selected these sites because they have similar climatic conditions. We made the assumption that nutrient dynamics in the six watersheds should be similar to nutrient dynamics in this portion of the Northern Piedmont ecoregion.

Frequency Distribution Based Approach

For frequency distribution based approach, we identified water quality samples collected by a variety of agencies from streams in the northern piedmont ecoregion stored in a variety of databases including the USEPA STORET and EMAP databases, United State Geological Survey (USGS) National Water Inventory System (NWIS) and National Water Quality Assessment (NAWQA) program, and the Maryland Biological Stream Survey (MBSS) database (Figure 2). Two populations of sites were developed. The first was all sites for which nutrient samples were available (All Sites). The second was all sites for which watershed land cover was available and for which reference criteria could be applied (Reference Sites).

The All Sites population included samples from all of the agencies in Table 1. For sites with multiple samples, samples were averaged to estimate an average site nutrient concentration. This reduced the influence of any one site on the percentiles. After all the

sites were prepared, we calculated the 25th percentile nutrient concentration of total phosphorus (TP) and total nitrogen (TN).

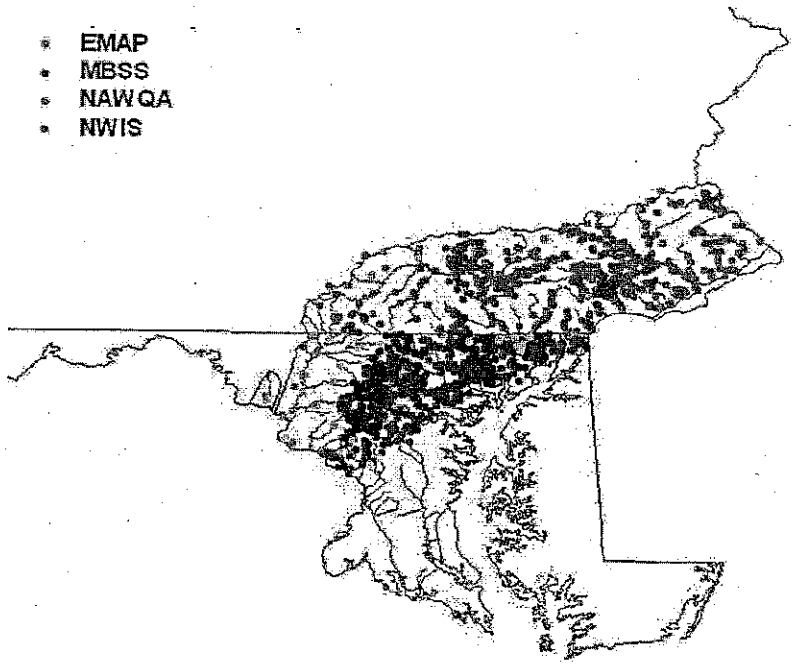


Figure 2 – Map of the sample sites used in the development of nutrient endpoints using the distribution based approach in this study, labeled by agency affiliation.

For sites where land cover information was available (USEPA EMAP, USGS NAWQA, and MBSS), we developed land cover screening criteria to identify least disturbed watersheds (sensu Stoddard et al. 2006). Least disturbed sites represent those watersheds with minimal human disturbance and, therefore, provide the best empirical estimate of chemical integrity. We developed two different reference criteria: >80% Forest, <5% urban (N=7) and >70% Forest, <5% urban (N=24). We then calculated the 75th percentile of total phosphorus and total nitrogen concentrations associated with these populations.

The results of the distribution based analyses gave comparable results whether the All Sites or either Reference Site population was used (Figure 3, Table 1). Total phosphorus concentrations were between 16 and 17 µg/L and total nitrogen concentrations between 1.3 and 1.5 mg/L (Table 1).

Table 1 – Values of TN and TP candidate endpoints derived using the distribution based approach.

	Reference Sites		All Sites
Parameter	>80% Forest <5% Urban	>70% Forest <5% Urban	
	75th Percentile	75th Percentile	25th Percentile
TN (mg/L)	1.5	1.3	1.5
TP (µg/L)	17	16	17
N	7	24	782 (TN) 836 (TP)

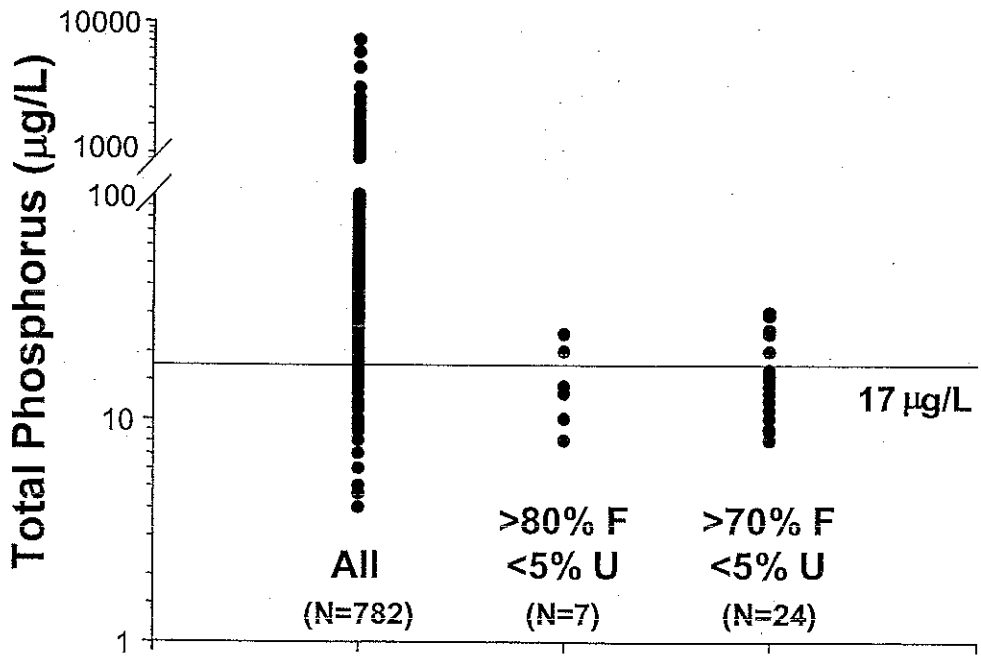


Figure 3 – Plot of total phosphorus samples in the All Sites and two Reference Site populations used to estimate candidate endpoints with the distribution based approach. Sample sizes are shown and the 25th percentile of All Sites and 75th percentile of Reference Sites was equal (17 µg/L).

Modeled Reference Expectation Approach

Another approach that falls under the rubric of "reference approaches" is the modeled reference expectation approach (Dodds and Oakes 2004). In this approach, multiple regression models of total nutrients versus human land cover (agriculture and urbanization) are built and then solved for the condition of no human land cover (i.e., the intercept). This approach has been used to estimate nutrient concentrations in the absence of human disturbance in the Midwest (Dodds and Oakes 2004).

We developed modeled reference expectation models for the northern piedmont region using data from the MBSS, USGS NAWQA and USEPA EMAP programs. The final equation for total nitrogen was:

$$\mathrm{Log}_{10}(\mathrm{TN}+1)=0.1+0.49(\mathrm{arcsine}\sqrt{\%\ \mathrm{Agriculture}})+0.14(\mathrm{arcsine}\sqrt{\%\ \mathrm{Urban}});$$

($R^2 = 0.43$, $F=125$, $p<0.001$).

Solving for the undisturbed condition leads to a modeled reference total nitrogen concentration of 260 µg/L.

No significant model for total phosphorus could be created with the land cover data, so we estimated the TP value for this approach based on N:P ratios.

N:P Ratios Suggest P Limitation Dominates the Northern Piedmont Ecoregion

We calculated N:P ratios for two populations of sites: All Sites in the northern piedmont dataset and Reference Sites in the northern piedmont dataset (Table 2). The average molar N:P ratio for All Sites was 259:1 and for Reference Sites was 184:1 or 208:1, depending on which reference criteria were used. We applied these ratios to the TN value estimated from the modeled reference expectation value for TN, which yielded TP values of 2, 3 and 3 µg/L TP, respectively. The molar ratio of N:P based on the

recommended USEPA nutrient criteria for this ecoregion (TP=36 µg/L, TN=690 µg/L) is 43:1. Applying this value, as well as the Redfield molar N:P ratio (16:1), to the value of TN estimated using the modeled reference expectation approach above led to estimated TP values of 14 and 37 µg/L, respectively. We would defend the use of natural ratios rather than Redfield given uncertainties in the applicability of Redfield to freshwater systems combined with the fact that Northern Piedmont average N:P ratios are much higher than Redfield. The 5th percentile of N:P ratios across all sites was 17 – meaning 95% of the streams sampled in the region have values above Redfield.

Table 2 – Values for molar N:P ratios estimated from the All Sites and the two Reference Site populations developed for use in this study. Molar N:P ratios were calculated as the ratios of moles TN: moles TP for each site.

	Reference Sites		All Sites
Parameter	>80% Forest <5% Urban	>70% Forest <5% Urban	
	N:P	N:P	N:P
Average	184	208	259
Median	186	186	158
25th Percentile	181	159	57
10th Percentile	141	87	25
5th Percentile	111	81	17

Stressor-Response Approach

Stressor-response approaches refer to a suite of analytical techniques that derive candidate endpoints by exploring the relationships between response variables and nutrient concentrations. Typical response variables in the context of nutrient endpoint development include water chemical aquatic life use indicators (dissolved oxygen, pH, etc.), algal biomass and/or algal assemblage metrics (e.g., percent nutrient sensitive diatoms), and aquatic life use indicators or biocriteria indicators (e.g., algal multimetric indices or individual metrics scores, invertebrate multimetric indices or individual

metrics, etc.). The value of these indicators is their direct linkage to aquatic life use designations. They, therefore, provide a way to connect nutrient concentrations directly to aquatic life use protection. We used a few different stressor-response analytical techniques to develop candidate nutrient endpoints using algal and invertebrate response indicators.

We selected two important nutrient variables to examine biological responses: total nitrogen (TN) and total phosphorus (TP). TN and TP are two of the four primary variables EPA recommended for nutrient endpoint development and are likely to limit aquatic primary producers. TP and TN may reflect stream trophic status better than inorganic P and N because nutrient depletion can be partially offset by increases in particulate fractions of TP and TN resulting from benthic algal drift and suspension in the water column (Dodds 2002). In addition, TN and TP are also measured more frequently in most of the national and state programs than other nutrient variables.

The primary response variable of interest for stream trophic state characterization is algal biomass, which is most commonly reported as mg m^{-2} Chl a. Chl a is a photosynthetic pigment and is a sensitive indicator of algal biomass. It is considered an important biological response variable for nutrient-related problems (USEPA 2000a). Periphyton is also often analyzed for dry mass (DM) and ash free dry mass (AFDM), which includes non-algal organisms. EPA also recommends a measure of turbidity as the response variable. However, turbidity is often associated with total suspended solids (TSS) and other environmental factors and is less commonly used as a direct response variable. In addition to these, algal species composition often responds dramatically to excess nutrients, including the proliferation of eutrophic and nuisance algal taxa. As a

result, algal metrics are frequently used as direct indicators of nutrient enrichment (van Dam et al. 1994, Pan et al. 1996). The last response variable we considered was macroinvertebrate metrics from multimetric indices. Macroinvertebrate indices are the most reliable and frequently used bioindicators, and many macroinvertebrate metrics are sensitive to nutrient enrichment.

Data:

We collected data from seven different national and state programs, similar to those used in the distribution based analyses (Table 3):

- USEPA Environmental Monitoring and Assessment Program (EMAP)
- United States Geological Survey (USGS) National Water-Quality Assessment (NAWQA) program
- USGS National Water Information System (NWIS)
- USEPA STORET database
- EPA national nutrient center (NNC) database (include Legacy STORET data)
- Maryland Biological Stream Survey (MBSS) program
- Pennsylvania Department of Environmental Protection (PADEP) periphyton biomass data

Two national projects, the USEPA EMAP and USGS NAWQA programs, simultaneously collected nutrients, periphyton, and macroinvertebrate composition data, which were valuable for exploring both algal and invertebrate assemblage responses to nutrients. The MBSS collected hundreds of macroinvertebrate samples from its statewide stream survey. This dataset was valuable for evaluating macroinvertebrate responses. Algal biomass data from NNC, PADEP, NWIS, and EMAP were used to evaluate nutrient algal biomass responses.

Table 3 - Biological data and their related chemical measurements in the stations in Northern Piedmont ecoregion. Numbers in parentheses are numbers of samples.

	EMAP	USGS NAWQA	USGS NWIS	STORET	MBSS	NNC	PADEP
TN	20	76	380		372	55	72
TP	20	76	347		372	54	72
Algal biomass	20		48			15	93
Algal species composition	20(47)	76(142)					
Macroinvertebrate composition	20(47)	77(106)		12(27)	658		

Data Analysis: Overview

Establishing definitive stressor-response relationships is a valuable line of evidence in the multiple lines of evidence approach. We first used Spearman correlation analysis to examine relationships between response and stressor variables. Correlation analyses identified significant relationships between biological response and nutrient variables. However, correlation may or may not indicate the real relationship. Numerous relationships were examined; only a subset of which was correlated. There were also results that were considered potentially important but showed weaker relationships (Appendix A).

We selected correlations of interest and performed visual scatter plots to further examine the relationships. We used either linear regression or a locally weighted average regression line to examine the trend of change along the environmental gradients. The locally weighted scatterplot smoothing (LOWESS) technique (Cleveland 1979) models nonlinear relationships where linear methods do not perform well. LOWESS fits simple models to localized subsets of the data to construct a function that describes, essentially,

the central tendency of the data. LOWESS fits segments of the data to the model. Tension, which describes the portion of data being used to fit each local function, was set at 0.50 for LOWESS regression.

We also used conditional probability analysis (Paul and MacDonald, 2005) to examine changes in the biological community along stressor gradients. Conditional probability provides the likelihood (probability) of a predefined response when a specific value of a pollutant stressor (condition) is exceeded. Conditional probability is the likelihood of an event when it is known that some other event has occurred. To estimate conditional probability of an impairment, we first had to define impairment as a specific value for a response variable (e.g., EPT < 8 genera). We used preexisting biocriteria thresholds as our response thresholds (MDNR 2005). Conditional probability answers the question: for a given threshold of a stressor, what is the cumulative probability of impairment? For example, if the total phosphorus concentration is greater than 30 μg/L, what is the probability of biological impairment (defined as < 8 EPT Taxa) for each site under consideration? All observed stressor values (in this example, all observed values of total phosphorous) are used to develop a curve of conditional probability (Paul and MacDonald, 2005). Because of its ability to identify risks of impact associated with given nutrient concentrations, the approach is suited to identifying nutrient thresholds protective of aquatic biological condition.

We also used nonparametric deviance reduction (change point analysis) to identify thresholds in biological responses to nutrients (Qian et al. 2003). This technique is similar to regression tree models, which are used to generate predictive models of response variables for one or more predictors. The change-point, in our application, is the first split

of a tree model with a single predictor variable (nutrient concentration). The loss function of regression trees can be evaluated by the proportion of reduction in error (PRE), which is analagous to the multiple R^2 of general linear models.

Data Analysis: Metric Calculation

Algal Metrics

A number of algal metrics were calculated to evaluate algal assemblage characteristics and their response to nutrient enrichment. Algal density, total algal species richness, diatom species richness, Shannon's Diversity index and evenness were calculated to measure abundance and diversity. The following diatom autoecological indices were calculated to characterize different specific algal assemblage responses:

Nitrogen uptake metabolism index: This index relates to the flow of nutrients, particularly nitrogen, within a waterbody. It is based upon the nitrogen cycling rate from autotrophs to heterotrophs to provide a measure of the nutrient input and processing occurring in a waterbody. The nitrogen uptake metabolism index (Van Dam et al. 1994) increases with elevated concentrations of organically bound nitrogen.

Saprobity index. This is a pollution tolerance index for algal species related to oxygen, organic matter, products of septic decay, and products of mineralization. The density of oligosaprobous diatoms and polysaprobous diatoms (Van Dam et al. 1994) is impacted by waters where oxygen is saturated or absent. The percentage of alpha-mesosaprobous to polysaprobous diatoms will increase as organic loads from human disturbance (e.g., from agriculture and wastewater discharges) increase.

Trophic state index (TSI). Eutraphentic diatoms (Van Dam et al. 1994) indicate elevated concentrations of nutrients that are important for diatom growth, including

nitrogen, phosphorus, inorganic carbon, and silica. As concentrations of these nutrients increase due to human disturbance, the TSI will also increase.

The TSI is one of the most important nutrient enrichment indices. Van Dam's TSI was developed in Europe and has been adapted in many parts of the United States. In addition, Jan Stevenson of Michigan State University, developed a TSI for the Mid-Atlantic highland region. He compiled a new TSI based on van Dam's TSI and Mid-Atlantic Highland (MAH) TSI into a new TSI. The Academy of Natural Science also developed a diatom TP inference model for the Northern Piedmont ecoregion. We calculated all the available TSI values to examine the relationship between observed TP concentrations and diatom inferred trophic states based on their indicator values and relative abundance. The percent sensitive taxa (indicator value from 1 to 2) were calculated (van Dam et al. 1994).

Oxygen requirement index. Several diatom taxa prefer conditions of high oxygen availability and will decrease in response to oxygen deficiency (Van Dam et al. 1994). Lower oxygen requirement index values implicate nutrient influx and subsequent eutrophication.

Macroinvertebrate Metrics

Approximately 40 macroinvertebrate assemblage metrics, including total taxa, Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa, Ephemeroptera taxa, percent Clingers, urban Intolerant %, percent Chironomidae, and Hilsenhoff's Biotic Index (HBI), were calculated using data from various programs. After initial screening, we selected a subset of benthic macroinvertebrate indicators, including the Maryland Benthic IBI score (B-IBI) and mayfly-stonefly-caddisfly (EPT) taxon richness. Other indicators

were all responsive general indicators of stress but were not diagnostic of any particular stressor.

Results: Algal Biomass – Nutrient Relationships

Similar to the distribution based approach, the molar N:P ratio of sites in the study region used for this analysis ranged from 5 to 2298 and averaged 259 and the 5th percentile of N:P ratios was greater than 16. As noted earlier, deviations from the Redfield ratio (41:7:1 by weight or 106:16:1 molar) are frequently used to determine N and P limitation (Redfield 1958). High N:P ratios indicate P is limiting growth, and low N:P ratios suggest that N is limiting growth. In addition to the strong evidence of P limitation from nutrient ratios, our examination of all the metrics with TN and other nitrogen parameters did not find strong correlations with biological variables. As a result, we considered Northern Piedmont streams as principally P-limited systems and focus on relationships with TP concentrations.

Not surprisingly, a strong algal biomass–nutrient relationship was not present in our examination of the datasets (Figure 4). A reasonable wedge shaped relationship was found between algal biomass and TP in the USGS NWIS dataset (black stars). The wedge shaped relationship is often found in large scale investigations when multiple stressors/constraints are present. It is possible that at some of the high nutrient concentration sites there was a light and flow limited accumulation of algal biomass. The wedge shaped relationship also indicated that elevated levels of algal biomass can exist at relatively low nutrient concentrations (<100 μg/L).

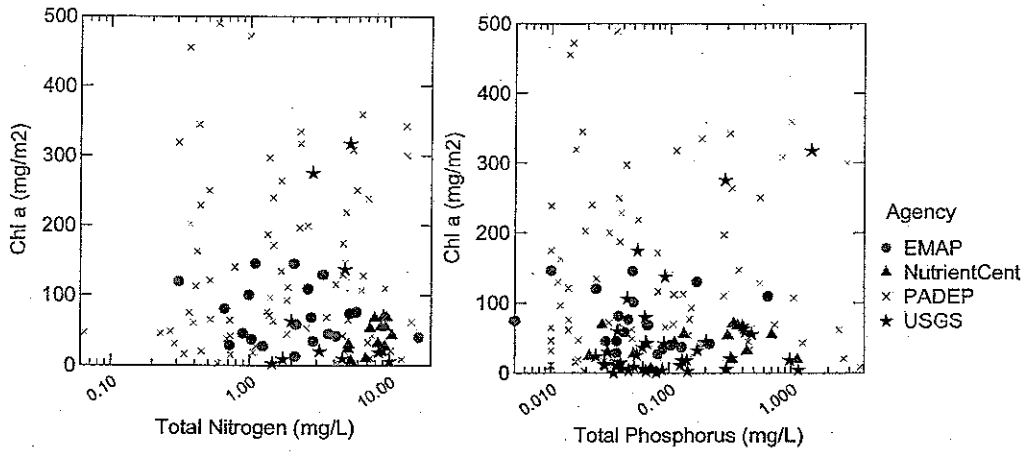


Figure 4 - Site average Chl a concentrations in relation to average TN and TP concentrations in the Northern Piedmont ecoregion. Data were collected by four different programs.

The samples with the highest algal biomass were collected by the PADEP - Pennsylvania State University periphyton study, which focused on the targeted watersheds. Surprisingly, the highest algal biomass occurred at sites where TP concentrations were relatively low (14–35 μg/L). It is possible that algal growth has been saturated even at this low level. The difference in magnitude of algal biomass among programs is probably due to different protocols being used by different programs.

Results: Algal Metrics – Nutrient Relationships

Algal compositional data from the EMAP and NAWQA programs were combined to obtain larger sample size. Two types of samples were collected and analyzed: depositional habitat and riffle habitat. Overall, four nutrient based metrics were significantly related to TP concentrations (Table 4).

Table 4 - Spearman Correlation matrices between environmental parameters and algal metrics. Significant Correlations are highlighted. (PANS = Philadelphia Academy of Natural Sciences, TSI = Trophic State Index, MSU = Michigan State University, MAIA = Mid-Atlantic Integrated Assessment)

	TN	*TP*	*NH4*	*NO3*	*DO*	*PH*	*COND*
Algal density	0.076	0.166	-0.023	0.064	-0.083	-0.083	0.147
Total taxa	0.004	0.236	0.279	-0.04	0.063	-0.131	0.133
Algal biovolume	-0.16	0.031	0.042	-0.131	0.035	-0.137	0.15
Diatom taxa	-0.03	0.243	0.283	-0.065	0.124	-0.127	0.12
N uptake index	-0.099	0.299	0.18	-0.13	0.267	-0.119	0.142
Oxygen index	-0.169	0.235	0.251	-0.184	0.255	-0.131	-0.006
Saprobity index	-0.142	0.208	0.229	-0.177	0.184	-0.127	-0.003
TSI index	-0.093	**0.402**	0.318	-0.107	0.243	-0.195	0.158
PANS TP model	-0.037	**0.515**	0.23	-0.037	0.072	0.163	**0.518**
MSU TSI index	-0.029	**0.505**	0.341	-0.039	0.188	-0.117	0.274
MSU-MAIA TSI index	-0.009	**0.454**	0.312	-0.063	0.17	-0.228	0.256

The Michigan State University (MSU) TSI index had the strongest correlation with TP concentrations among the three TSI indices. The diatom TP inference index from the Philadelphia Academy of Natural Science was the best predictor of TP concentrations in the Northern Piedmont ecoregion. However, since the model was developed from the same dataset, we did not use it for further analysis.

MSU TSI was plotted against TP concentrations in both riffle and depositional samples (Figure 5a and 5c). The significant relationship between TSI and TP concentrations in both sample types supports the prediction that TP is associated with increased trophic state in these streams. The relatively small variance explained by the regression models (R^2=0.22 and 0.35 respectively) is likely due to other stressors coexisting in the streams that are affecting diatom species compositions. A conditional probability analysis was performed for each of these sample types (Figures 5b and 5d). This analysis identifies TP thresholds associated with an increase in the probability of adverse ecological conditions, in this case, a shift in the TSI from meso- to eutraphentic conditions. According to van Dam et al (1994), TSI=4 indicates a meso-eutraphentic

condition and 5 indicates eutraphentic condition. We defined the adverse condition as TSI=4.5, which is the transition from meso-eutraphentic to eutraphentic. The conditional probability analyses indicates that the probability of impairment (TSI>4,5) increases with elevated TP concentrations, i.e., diatom species composition shifts from meso-eutraphentic taxa to eutraphentic taxa. The threshold associated with the change point of this relationship was 36 μg/L for both riffle and depositional samples.

Results: Benthic Macroinvertebrate Metrics – Nutrient Relationships

The largest dataset available for analyzing macroinvertebrate responses to nutrient concentrations was the Maryland Biological Stream Survey (MBSS) dataset. The MBSS sampled more than 500 macroinvertebrate samples with 372 corresponding nutrient samples in the Northern Piedmont ecoregion. The MBSS also developed a benthic macroinvertebrate IBI for this ecoregion based on the scores of six metrics: total taxa, Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa, Ephemeroptera taxa, % Clingers, % Intolerant Urban, % Chironomidae. For each metric, scoring criteria were developed based on the distribution of values from least disturbed reference sites (score of 5). We selected the middle point of the distribution as the impairment threshold for each metric (Table 5).

Of the six metrics above, only EPT taxa (Pearson $r= -0.293$, $p<0.001$), %intolerant urban taxa ($r=-193$, $p=0.005$), and % clingers ($r=0.191$, $p=0.006$) were negatively correlated with log TP concentrations. The other three metrics were either not sensitive to nutrient enrichment or more sensitive to other stressors.

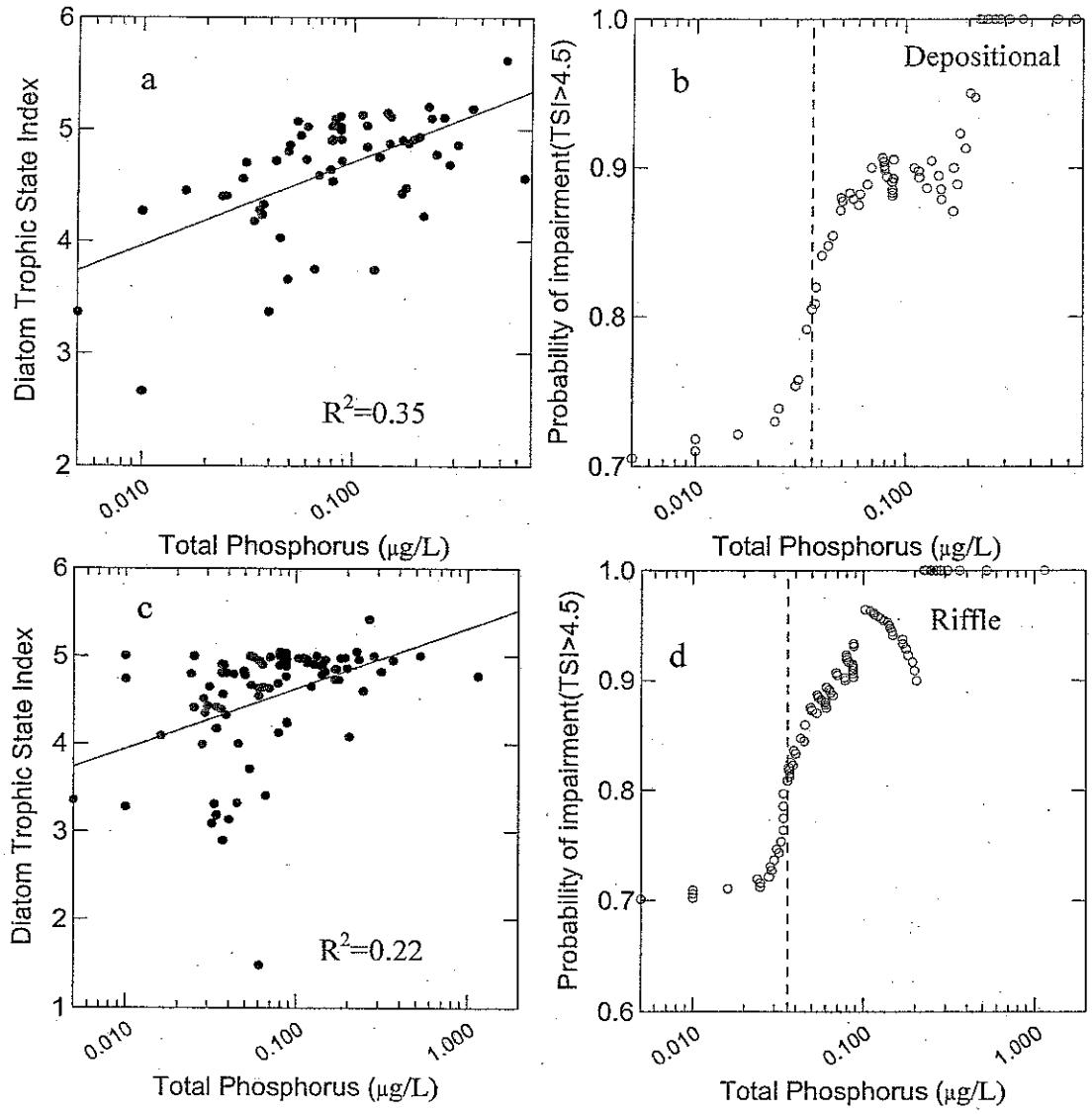


Figure 5 - Response of diatom trophic state index (MSU_TSI) to total phosphorus concentrations in depositional habitats and riffles. The figures on the right show the conditional probability of exceeding a TSI value of 4.5 with increased TP concentration.

Table 5 – Threshold values for the MBSS benthic macroinvertebrate IBI metrics in the Northern Piedmont ecoregion.

Scoring criteria	5	3	1	Mid Point
Number of Taxa	≥ 25	15 – 24	<15	20
Number of EPT	≥ 11	5 – 10	< 5	8
Number of *Ephemeroptera*	≥ 4	4– 3	< 2	3
% Chironomidae	≤ 4.6	4.7 – 63	> 63	38.6%
% Intolerant Urban	≥ 51	12 – 50	< 12	31.5%
% Clingers	≥ 74	31 – 73	< 31	52.5%

For example, EPT taxa declined with increased TP concentrations (Figure 6a). The scatterplot relationship exhibited a traditional wedge shape decline, while the conditional probability graph (Figure 6b) clearly indicated that the probability of impairment (EPT taxa<8) increased from 45 to 88% when TP concentrations increased from 30 to 80 μg/L TP. Change point analysis indicated a threshold at 38 μg/L.

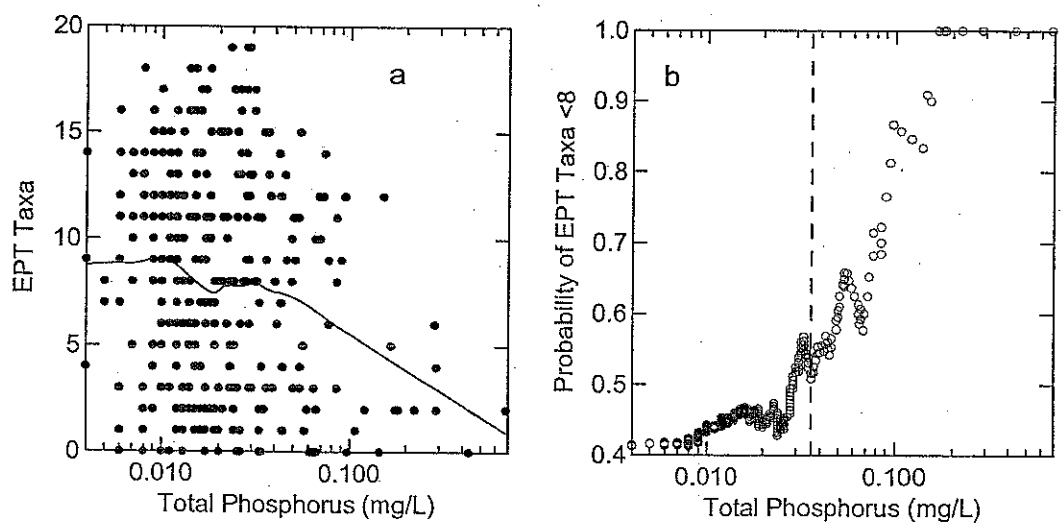


Figure 6 - Response of EPT taxa metric to the phosphorus gradient in the Northern Piedmont Ecoregion. Figure on the right shows the conditional probability of having fewer than 8 EPT taxa as TP concentrations increase.

The percent of clinger taxa in the streams responded similarly to increased TP concentrations (Figure 7). The conditional probability graph (Figure 7b) showed that the probability of impairment (% Clinger taxa <52.5) would likely increase from 45 to 70% when TP increased from 30 to 80 µg/L. Change point analysis identified a threshold at 39 µg/L.

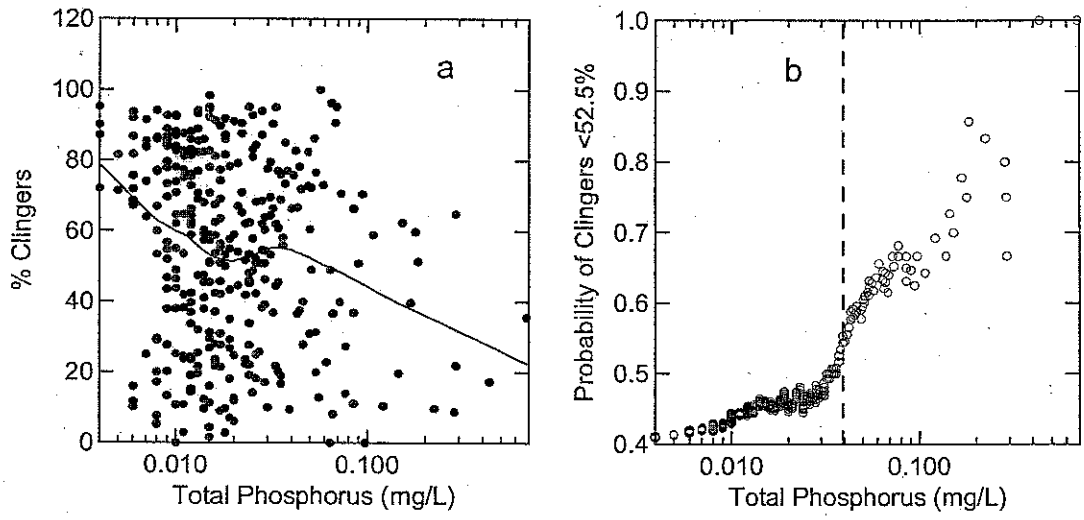


Figure 7 - Response of % clingers in benthic samples along phosphorus gradient in the Northern Piedmont Ecoregion. The figure on the right shows the conditional probability of having fewer than 52.5% clinger taxa in a stream as TP concentration increases.

The % urban intolerant taxa metric was based on the response of taxa intolerant of urbanization (Figure 8). Although TP concentration likely increases along the urban gradient, this metric is less sensitive to nutrient concentrations since it includes some taxa that are insensitive to nutrients but sensitive to other stressors. The conditional probability graph (Figure 8b) and change-point analysis indicated that probability of impairment would likely increase from 50% to much higher after TP concentration exceeded 64 µg/L TP.

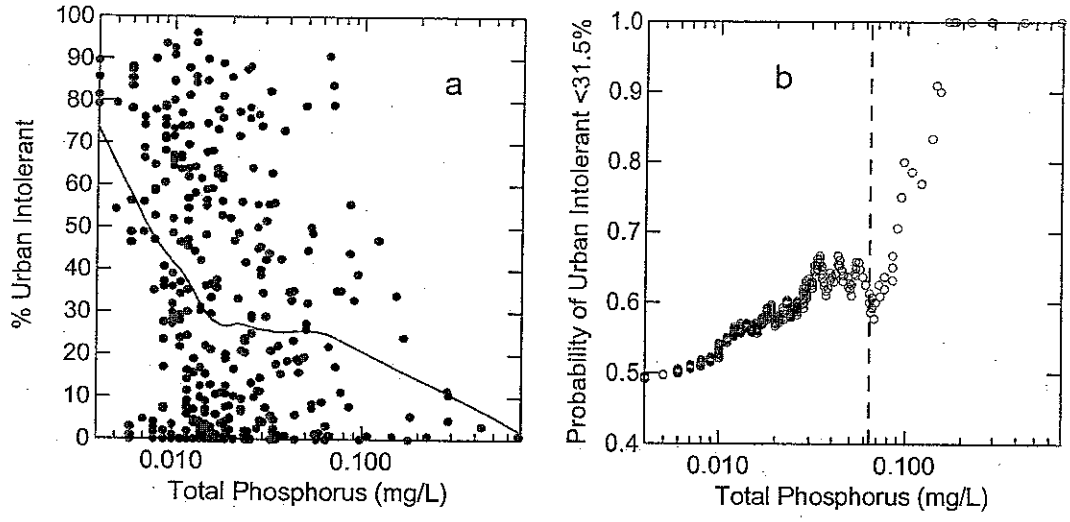


Figure 8 - Response of % intolerant urban taxa in benthic samples along phosphorus gradient in the Northern Piedmont Ecoregion. The figure on the right shows the conditional probability of having fewer than 31.5% urban intolerant taxa in a stream as TP concentration increases.

Responses of macroinvertebrate metrics to nutrient enrichments have been examined for other databases. The Spearman correlation matrices are included in Appendix 1 and 2. Due to relatively small sample size and a lack of relationships of interest, we did not examine these further.

Literature Based Analysis: Current Existing Endpoints or Threshold Values

In this last analytical section, we present several studies relevant to the development of nutrient endpoints in the northern Piedmont region of Pennsylvania. These are taken principally from the peer-reviewed literature and reflect increasing experimental and theoretical interest in the impact of nutrients on natural stream systems. We attempted to extract information from these studies that could recommend specific endpoints.

In natural, shaded streams [such as those evaluated in the Dodds et al. (2002) model], it is difficult to assess the full growth potential of algae. Algal growth potential has been

evaluated using artificial stream channels that are fully exposed to nutrient and light gradients. Previous studies (Horner et al. 1983, Bothwell 1989) demonstrated that in artificial streams, algal growth could be saturated (i.e., achieved maximum growth rate) at 25–50 μg/l phosphorus. Rier and Stevenson (2006) found that at 16 μg/L soluble reactive phosphorus (SRP) or 86 μg/L dissolved inorganic nitrogen (DIN), algal growth was at 90% of its maximum rate. They also found that saturation concentrations were 3–5 times lower than concentrations needed to produce maximum algal biomass (i.e., 430 μg/L DIN and 80 μg/L SRP for growth saturation). However, these values were derived mostly on the basis of diatom and bluegreen algae growth. We expect that green algae (i.e., *Cladophora*) would have higher nutrient saturation concentrations for peak growth (Borchardt 1996).

Studies in adjacent regions have shown consistently low values for TP required for the control of benthic chlorophyll. An ongoing study on the Jackson River in Virginia (Louis Berger Group, pers. comm) is proposing an ortho-phosphorus endpoint of 38 μg/L. This is based on a regression equation developed using local data. In New Jersey, a trophic diatom index (TDI) was developed that identified a TP below 25 μg/L as a low, protective TDI value and a range from 75 μg/L to 100 μg/L for a high TDI value (Ponader et al. 2005). These authors also found that within this region, concentrations above 50 μg/L TP were sufficient to produce nuisance algal growth.

EPA's nutrient threshold recommendation for the Northern Piedmont was 690 μg/L for TN and 37 μg/L for TP. Ponader and Charles (2003) applied EPA's reference approach to the Northern Piedmont in New Jersey and found fairly good agreement with

the EPA recommended numbers. The Ponader and Charles (2003) estimates using the distribution approach were 1.3 μg/L for TN and 40-51 μg/L for TP.

Dodds and Welch (2000) conducted a meta-study including values from a range of areas nationwide. These were combined into regression equations to predict chlorophyll. They found that if a mean of 50 mg m^{-2} of chlorophyll is the target (thus insuring chlorophyll is less than 100 mg m^{-2} most of the time), TN should be 470 μg/L and TP should be 60 μg/L. Even lower numbers should be considered for more pristine waters. These estimates were more general in scope. These authors further noted that lower TN and TP values associated with these chlorophyll concentrations were obtained when using a detailed, smaller data set than those from a larger data set (55 μg/L TP from a large dataset versus 21 μg/L for a more specific, local data set).

USGS conducted a study in 2001 for a broad area of the US, including the New River and Big Sandy River in Virginia (Robertson et al. 2001). They looked at 234 sites using the reference approach and found that a TP of 20 μg/L was appropriate for what they define as Environmental Nutrient Zone 2. Similarly, Ponader et al. (2005) in a study of over 35 streams in Virginia observed changes in the diatom assemblages and suggested threshold limits of 500 μg/L for TN and 50 μg/L for TP to protect against conditions that they termed as "nutrient impaired", based on a variety of factors.

ENSR (2003) developed endpoints for rivers in New England by combining the distribution based approach (using a database of 569 stream and river nutrient and trophic parameter data) and effects-based approach (based on a weight-of-evidence review of literature, models and TMDL studies)(Table 6). Using the EPA approach for calculating ambient water quality recommendations, the 25th percentile of all rivers and streams and

the 75th percentile of the reference waters provided relatively similar values. ENSR (2003) suggested, based on the weight-of-evidence, that 40 µg/L TP and 800 µg/L TN would be an upper bound for nutrient endpoints (i.e., approaching impaired aquatic community status).

Table 6 - Comparison of New England Water Quality Recommendations, after ENSR (2003) with EPA recommended regional criteria. All values in µg/L.

		New England Ecoregion All Season		EPA Recommended Criteria
Sub-Ecoregions		25th Percentile	75th Percentile	
83	Chl a	1.6	1.5	1.6
	TN	470	538	480
	TP	31	44	24.1
82	Chl a	1.7	2.5	
	TN	33	325	390
	TP	14	12	12
59	Chl a	4.9		
	TN	560	458	570
	TP	20	22	23.5
58	Chl a	2.2		3.4
	TN	360	121	420
	TP	10	12	5
Composite	Chl a	1.9	1.8	
	TN	460	520	
	TP	20	23	

Rohm et al. (2002) conducted a national study to demonstrate how regional reference conditions and draft nutrient endpoints could be developed. They divided the country into 14 regions and analyzed available nutrient data as a case study, using EMAP data from Central and Eastern Forested Uplands, an area that includes much of central Pennsylvania. This case study suggested a criterion of 375 µg/L for TN and 13 µg/L for TP. Rough estimates from the data presented for their Region IX that includes Eastern Pennsylvania gives estimates of 500 µg/L TN and 20 µg/L TP.

Several states have developed nutrient standards or guidelines (see companion report). These values range from a maximum TP of 100 µg/L to a summer average TP of 25 µg/L. Delaware uses TP in assessing their waters for reporting under Section 305(b) of the Clean Water Act. They list segments as impaired when one or more stations whose lower confidence limit is at or above the moderate value of 1.0 to 3.0 mg/L TN and 50 to 100 µg/l TP. Hill and Devlin (2003), in a preliminary study in Virginia, suggest a TN threshold for benthic impairment somewhere between 350 and 900 µg/L TN. The Dodds et al. (2006) TN breakpoint fits within this range for TN.

Recommended Endpoints

Total Phosphorus (TP)

Endpoint (magnitude)

Our analyses relied on a weight-of-evidence analysis drawing on many different analytical approaches. Each of the different approaches produced slightly different endpoints and these are summarized in Table 7.

In a weight-of-evidence approach, the different analyses are weighted based, essentially, on their applicability and the strength of the analysis. The Reference Approach analyses we weight less for a few reasons. The Reference Site 75th percentile estimate was based on few sites and the All Sites 25th percentile analysis included a variety of data spanning many different periods. The modeled reference expectation approach could not produce a significant model for TP and the TP endpoint for this method was derived, instead, from the TN model using an estimate based on a variety of TN:TP molar ratios. Lastly, the reference approach is less easy to link directly to use

protection, given that it is based on percentiles of a frequency distribution. Reference sites arguably reflect the "indigenous" or "natural" condition, which is the goal of aquatic life use standards in Pennsylvania, but this is an indirect measure.

Table 7 – Summary of candidate endpoints for each of the analytical approaches discussed.

	Approach	TP Endpoint (µg/L)
Reference Approach		**2-37**
	Reference Site 75th Percentile	16-17
	All Sites 25th Percentile	17
	Modeled Reference Expectation	2-37
Stressor-Response		**36-64**
	Conditional Probability – EPT taxa	38
	Conditional Probability - % Clingers	39
	Conditional Probability - % Urban Intolerant	64
	Conditional Probability - Diatoms TSI	36
Other Literature		**13-100**
	USEPA Recommended Regional Criteria	37
	USEPA Regional Criteria Approach – Local Data	40-51
	Algal Growth Saturation	25-50
	Nationwide Meta-Study TP-Chlorophyll	21-60
	USGS Regional Reference Study	20
	USGS National Nutrient Criteria Study	13-20
	New England Nutrient Criteria Study	40
	Virginia Nutrient Criteria Study	50
	New Jersey TDI	25-50
	Delaware Criteria	50-100

The stressor-response analyses carry more weight because we could link nutrient concentrations to specific aquatic life endpoints – both invertebrate and algal. Using invertebrate taxa metrics, conditional probability analyses evaluated those TP concentrations which increased the risk of exceeding degradation thresholds developed for these macroinvertebrate metrics in comparable piedmont streams in Maryland. For the diatom Trophic State Index (TSI), the same analysis was used to identify the TP concentration associated with a shift from meso- to eutrophic conditions.

The scientific literature was variably weighted, since it included data from regions proximate to Pennsylvania, as well as data less applicable to Pennsylvania.

Based on greater weight to the stressor-response models, we recommend TP endpoints for northern piedmont streams in southeastern Pennsylvania of 40 μg/L. This value is comparable to the majority of the stressor-response analyses, on the high end of the reference approaches, and intermediate to the scientific literature values, but comparable to regionally relevant literature values.

• Recommended endpoint: 40 μg/L TP

Sample period

We recommend applying the endpoint over the algal growing season (April to October), which in streams is typically the time during which the greatest risk of deleterious algal growth exists.

• Endpoint applies from April to October

Sample duration

Unlike toxics, there is less literature to recommend appropriate sample duration and frequencies for nutrients. Toxics, with chronic and acute criteria, have a longer history of implementation. Their mode of action is also very different than nutrients. As a result, it was more difficult to recommend an appropriate sample period than to derive the endpoints themselves.

Humans tend to sample nutrients at temporal scales that are different than those to which stream organisms respond. Streams respond both to pulsed as well as chronic nutrient concentrations. For example, algae possess mechanisms to store nutrients and use these stored nutrients for growth over time – so they can respond to episodic inputs.

Moreover, the responses to episodic inputs include both assemblage responses (for example, development the nuisance algal taxa) as well as population and individual responses (biomass).

The nutrient data we analyzed for the invertebrate and plant responses were based on single grab samples associated with biological sampling. These analyses, therefore, represent a space for time substitution of sorts, estimating what would occur in a piedmont stream as nutrient concentrations increase.

• Endpoint is assessed as the average TP concentration during the growing period over one year.

These factors would recommend a not-to-exceed criterion. However, water velocity affects nutrient delivery in streams and elevated nutrients associated with high flows may not be as accessible to benthic algae. We also recognize that there is resistance to not-to-exceed standards and concern about the risk of capturing false positives, even though the risk of false negatives is similarly great. These concerns would recommend averaging multiple samples over some time period. Algal and microbial responses to nutrients can occur rapidly, but these can be offset by floods that scour the bottom and remove algae. At this time, there is limited information and we have had insufficient time to investigate appropriate averaging periods, especially those that result in conditions detrimental to uses.

As a result, for the purposes of these TMDLs, we recommend that the TP endpoint be applied as an average of water samples taken over the growing season. Realize, again, that there is less information to guide this recommendation, which is based principally on our professional judgment and in an attempt to be consistent with other typical duration

procedures. A more conservative alternative would be to use the recommended endpoint as a not-to-exceed value, but again, we have had insufficient time to evaluate this.

We feel that this approach will be protective, but we strongly encourage the state and EPA to investigate this issue more fully for the purposes of regional criteria development. For the TMDLs, this approach is sufficient, but it deserves more attention and resources before being applied to regional criteria.

Total Nitrogen (TN)

Endpoint

The focus of our work was the development of total phosphorus endpoints. This was principally due to the fact that TP was assessed as the cause of impairment. Our analyses support the conclusion that these streams are P limited, based on instream N:P molar ratios evaluated against Redfield. The distributional statistics of N:P ratios taken from more than 552 stream sites across the northern piedmont region in Pennsylvania and Maryland are shown in Table 2.

The traditional, critical N:P Redfield molar ratio is 16:1, values below indicating N limitation and those above, P limitation. Ratios have to be considered in relation to supply and become less meaningful as nutrient supplies exceed uptake capacity of streams. Even so, clearly more than 95 percent of the streams sampled in the northern piedmont region were P limited, relative to Redfield.

Because these systems are not N limited, relationships between TN and response measures are less well etablished. The fact that N is not limiting also means that TN likely contributes less to use impairment from eutrophication in this region. Endpoints are best derived when clear connections to use impairment can be made. It is most likely

that N contributes to use issues in the tidal and estuarine waterbodies downstream of rivers and streams in this region (e.g., Delaware Bay). Those systems are where data and analyses will be able to suggest an appropriate N target for upstream systems. That being the case, there is some risk in setting stream TN endpoints in this region that may ultimately be inconsistent with those needed to protect uses from TN enrichment in the Bay. While we cannot recommend an N target, we can recommend a few TN endpoints using different approaches:

Ratio-based

This approach assumes it is protective to reduce N in proportion to P based on ambient molar nutrient ratios. This, again, may not sufficiently reduce N to protect downstream uses, but it would keep the ratios consistent.

Based on the distributional statistics on stream molar N:P ratios (Table 2), one could decide that keeping the ratio consistent with average Piedmont streams would be appropriate. The average piedmont TN:TP molar ratio for reference streams (minimal human disturbance using 70% forest cover and 5% urban cover) was 208:1. Based on a 40 ug/L TP target, that would recommend a TN value of 3.8 mg/L TN. Note that this value would be consistent with other Piedmont streams, which is on the higher end for TN, especially for export to an estuary. One could also use the ratio based on EPA recommended criteria (43:1). Based on that ratio, the TN concentration would be 780 μg/L.

Modeled Reference Expectation

We ran an analysis of human disturbed land cover against TN and found a significant linear regression model (see above), where TN increased with non-forested land cover. The equation was:

$$\mathrm{Log}_{10}(\mathrm{TN}+1) = 0.1 + 0.49(\mathrm{asin}\sqrt{\%\ \mathrm{Agriculture}}) + 0.14(\mathrm{asin}\sqrt{\%\ \mathrm{Urban}})$$

Solving for the y-intercept (all forested land) leads to a value of LogTN = 0.1 or 260 µg/L. This would be the expected average TN concentration without human disturbance in the watershed, based on this modeling approach. Note that this value is much less than the ratio-based number, but is likely more along the lines of what might have existed in the absence of land-cover disturbance.

EPA Regional Recommendation

The EPA recommended criteria based on their reference population derived approach was 690 µg/L. That is another value that could be used.

Distribution Based

This approach looked at the distribution of TN values in reference sites and across all sites we gathered for this analysis, comparable to those used for the TP endpoint. The 75^{Th} percentile of the reference population values ranged from 1.3 to 1.5 mg/L TN based on what land cover criteria were used (Table 1). The value based on the 25th percentile of all sites would be 1.5 mg/L.

Overall

The endpoints ranged from 260 µg/L to 3.7 mg/L. Again, the less established linkage to use protection as seen for the TP endpoint stressor-response analysis, makes the selection difficult. A value approximate to 1.5 mg/L would seem to be sensibly

consistent with reference conditions (perhaps the strongest analysis in this context). This would be applied over the same time period and sample duration as TP.

Again, we cannot be as definitive as we were with TP, because there appears to be little reason to think that TN is limiting uses in these northern piedmont freshwater stream systems – but rather the effects of N are likely manifested downstream. It is those systems that should be driving the choice of protective TN targets upstream. Given that, the strategy would be to err on the low side, and this would argue for something more along the 1-1.5 mg/L range.

Literature Cited

Borchardt, M. A. 1996. Nutrients. Pages 183–227 in R. J. Stevenson, M. L. Bothwell, and R. L. Lowe, editors. Algal ecology: freshwater benthic ecosystems. Academic Press, San Diego.

Bothwell, M. L. 1989. Phosphorus Limited Growth Dynamics of Lotic Periphytic Diatom Communities - Areal Biomass and Cellular Growth-Rate Responses. Canadian Journal of Fisheries and Aquatic Sciences 8(46): 1293-1301.

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

Dodds, W.K. 2002. Freshwater Ecology: Concepts and Environmental Applications. Academic Press, San Diego.

Dodds, W. K. and R. M. Oakes,. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography-Methods (2): 333-341.

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

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

Dodds, W.K. and E.B. Welch. 2000. Establishing nutrient criteria in streams. Journal of the North American Benthological Society 19:186-196

ENSR, International. 2003. Collection and evaluation of ambient nutrient data for rivers and streams in New England. New England Interstate Water Pollution Control Commission, Technical Report. http://www.neiwpcc.org/ncreports.asp

Hill, J. and G. Devlin. 2003. Memorandum: Associations Between Biological, Habitat, and Ambient Data in Upland Western Virginia Ecoregions. Virginia Department of Environmental Quality, West Central Regional Office. Roanoke, VA. (December 17, 2003).

Horner, R.R., E.B. Welch, and R.B. Veenstra. 1983. Development of nuisance periphytic algae in laboratory streams in relation to enrichment and velocity. Pages 121-134 in R.G. Wetzel (ed.). Periphyton of Freshwater Ecosystems. Dr. W. Junk Publishers. The Hague.

Maryland Department of Natural Resources (MDNR). 2005. New Biological Indicators to Better Assess the Condition of Maryland Streams. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment, Annapolis, MD. CBWP-MANTA-EQ-05-13.

Pan, Y. D., R. J. Stevenson, B. H. Hill, A. T. Herlihy and G. B. Collins,. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. Journal of the North American Benthological Society 4(15): 481-495.

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

Ponader, K. & D. F. Charles. 2003. Understanding the relationship between natural conditions and loadings on eutrophication: Algal indicators of eutrophication for New Jersey streams. Final Report Year 2. Report No. 03-04. Patrick Center for Environmental Research, Academy of Natural Sciences. Philadelphia, PA.

Ponader, K., C. Flinders, and D. Charles. 2005. The Development of Algae-based Water Quality Monitoring Tools for Virginia Streams. Report No. 05-09 for the West Central Regional Office, Virginia Department of Environmental Quality. Patrick Center for Environmental Research, Academy of Natural Sciences. Philadelphia, PA.

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

Redfield, A.C. 1958. The biological control of chemical factors in the environment. American Scientist 46, 205-221.

Rier, S. T. and R. J. Stevenson. 2006. Response of periphytic algae to gradients in nitrogen and phosphorus in streamside mesocosms. Hydrobiologia 561:131-147.

Robertson, D.M., D.A. Saad, A.M. Wieben, U.S. Environmental Protection Agency Region V, U.S. Environmental Protection Agency Region VII, and U.S. Geological Survey. 2001. An Alternative Regionalization Scheme for Defining Nutrient Criteria for Rivers and Streams. U.S. Geological Survey Water-Resources Investigations Report 01-4073. Middleton, WI.

Rohm, C.M., J.M. Omernik, A.J. Woods, and J.L. Stoddard. 2002. Regional characteristics of nutrient concentrations in streams and their application to nutrient criteria development. Journal of the American Water Resources Association 38: 213-239.

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

United States Environmental Protection Agency (USEPA). 2000a. Nutrient Criteria Technical Guidance Manual. Rivers and Streams. United States Environmental Protection Agency, Office of Water, Washington, DC. EPA-822-B-00-002

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

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

United States Environmental Protection Agency (USEPA). 2007. Nutrient Criteria Technical Guidance Manual. Estuarine and Coastal Marine Waters. United States Environmental Protection Agency, Office of Water, Washington, DC. EPA-822-R-07-004

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

Appendix A. Spearman Correlation matrices among enrichment related environmental variables and a few macroinvertebrate metrics in Northern Piedmont Ecoregion in EMAP database. Significant correlations are highlighted.

	Conductivity	DOC	NH4	TN	TP
Total Richness	-0.22	-0.387	**-0.486**	-0.014	-0.134
Chironomid richness	-0.233	**-0.411**	-0.343	-0.043	-0.229
Ephemeroptera richness	-0.004	-0.157	**-0.431**	0.107	0.138
EPT richness	-0.177	-0.321	**-0.438**	0.005	-0.044
Non-insect richness	0.107	0.081	-0.078	0.181	0.054
Megaloptera richness	0.066	0.034	-0.112	-0.177	-0.261
Oligochaete/leech richness	**-0.417**	-0.153	-0.265	-0.32	-0.095
Plecoptera richness	**-0.405**	**-0.616**	-0.379	-0.114	-0.373
Trichoptera richness	-0.158	-0.269	-0.334	-0.005	-0.038
Collector-filterer richness	0.021	-0.256	-0.259	0.172	-0.04
Collector-gatherer richness	-0.325	**-0.43**	**-0.478**	-0.027	-0.187
Mixed functional richness	-0.241	**-0.537**	**-0.499**	0.091	-0.067
Omnivore richness	0.317	0.124	0.166	0.235	0.058
Predator richness	-0.154	-0.244	-0.177	-0.301	-0.3
Scavenger richness	-0.333	-0.091	-0.213	-0.217	-0.05
Shredder richness	-0.083	-0.306	**-0.489**	0.136	0.007
Scraper richness	0.056	-0.042	-0.289	-0.055	-0.147
Intolerant taxa richness	-0.381	**-0.508**	**-0.543**	-0.065	-0.201
Facilitator richness	-0.163	-0.372	**-0.429**	0.011	-0.19
Tolerant richness	0.083	0.138	-0.047	-0.05	0.084
Tolerant %	0.283	**0.449**	0.268	-0.017	0.23
Dominant Taxon %	0.204	0.246	0.182	-0.037	0.183
Shannon's diversity	-0.194	-0.387	-0.333	0.036	-0.22
Simpson's diversity	0.194	0.302	0.211	-0.058	0.203
Dominant 3 Taxa %	0.205	0.293	0.202	-0.088	0.201
Hilsenhoff BI	0.245	0.289	0.312	0.059	0.194
Sample Size	47	45	45	47	47

Appendix B. Spearman Correlation coefficients among enrichment related environmental variables and a sub-sample of macroinvertebrate metrics from Northern Piedmont Ecoregion stream samples in the USGS NAWQA database. Significant correlations are highlighted.

	Conductivity	NH4	TKN	NOX	TN	DP	TP
Beck's BI	**-0.586**	-0.147	-0.29	0.039	0.012	-0.238	-0.151
Bivalves %	0.274	0.21	**0.668**	-0.317	-0.254	**0.518**	0.205
Chironomid %	0.217	0.28	0.064	0.111	0.133	-0.033	-0.05
Chironomid richness	0.028	0.264	0.163	-0.009	0.036	-0.022	0.039
Collector %	0.143	-0.125	-0.164	0.289	0.388	-0.233	-0.24
Collector richness	-0.273	-0.038	-0.318	0.086	0.089	-0.3	-0.195
Clinger richness	**-0.551**	-0.04	-0.264	-0.08	-0.076	-0.15	-0.138
Coleoptera %	0.042	-0.07	0.346	0.137	0.115	0.415	**0.511**
Coleoptera richness	-0.417	0.148	-0.016	-0.126	-0.201	0.127	0.26
Percent Corbicula	0.471	0.434	0.401	-0.064	0.015	0.386	0.37
Crustacea+Mollusca %	0.157	0.409	0.459	-0.46	-0.429	0.424	0.286
Margalef Diversity	-0.467	-0.073	-0.294	-0.067	-0.045	-0.239	-0.14
Diptera richness	-0.065	0.247	0.135	0.07	0.098	-0.032	0.033
Dominant Taxon %	0.271	0.148	0.293	-0.054	-0.099	0.33	0.331
Ephemeroptera richness	-0.424	-0.274	-0.113	0.062	0.056	-0.036	-0.165
EPT richness	-0.47	-0.198	-0.289	0.036	0.006	-0.163	-0.139
Evenness	-0.278	-0.254	-0.387	0.174	0.217	-0.369	-0.214
Filterer %	-0.002	0.434	0.163	-0.168	-0.153	0.113	0.175
Filterer richness	-0.231	0.154	0.012	-0.252	-0.289	0.014	0.133
Hilsenhoff BI	0.225	0.032	-0.004	-0.066	-0.057	-0.203	0.055
Intolerant richness	**-0.577**	-0.156	-0.304	-0.006	-0.024	-0.272	-0.155
Noninsect %	0.171	-0.269	-0.042	-0.004	0.039	-0.112	0.03
Odonate %	-0.173	-0.104	-0.049	0.023	0.019	-0.032	-0.057
Oligochaete %	-0.009	-0.12	-0.159	-0.058	-0.062	-0.24	-0.208
Oligochaete richness	-0.115	-0.02	-0.284	0.012	0.041	-0.283	-0.225
Plecoptera %	**-0.525**	-0.229	-0.346	0.066	0.078	-0.417	-0.315
Plecoptera richness	**-0.533**	-0.173	-0.331	0.048	0.06	-0.399	-0.318
Predator %	-0.236	-0.336	-0.423	0.183	0.21	-0.493	-0.256
Predator richness	-0.424	-0.051	-0.065	-0.255	-0.252	-0.051	-0.159
Scraper Richness	-0.184	0.222	0.101	-0.417	-0.444	0.262	0.175
Shannon diversity	-0.397	-0.222	-0.423	-0.061	-0.011	-0.421	-0.324
Swimmer %	**-0.578**	-0.337	**-0.508**	-0.001	-0.023	-0.398	-0.392
Swimmer richness	**-0.564**	-0.362	-0.477	0.041	-0.004	-0.376	-0.297
Tolerant %	0.099	-0.098	-0.07	0.092	0.098	-0.158	0.114
Tolerant richness	0.258	0.281	0.344	-0.313	-0.367	0.322	0.251
Total richness	-0.441	0.074	-0.174	-0.194	-0.184	-0.106	-0.037
Trichoptera richness	-0.319	0.006	-0.264	-0.216	-0.243	-0.15	-0.047
Sample size	77	52	44	52	44	44	38

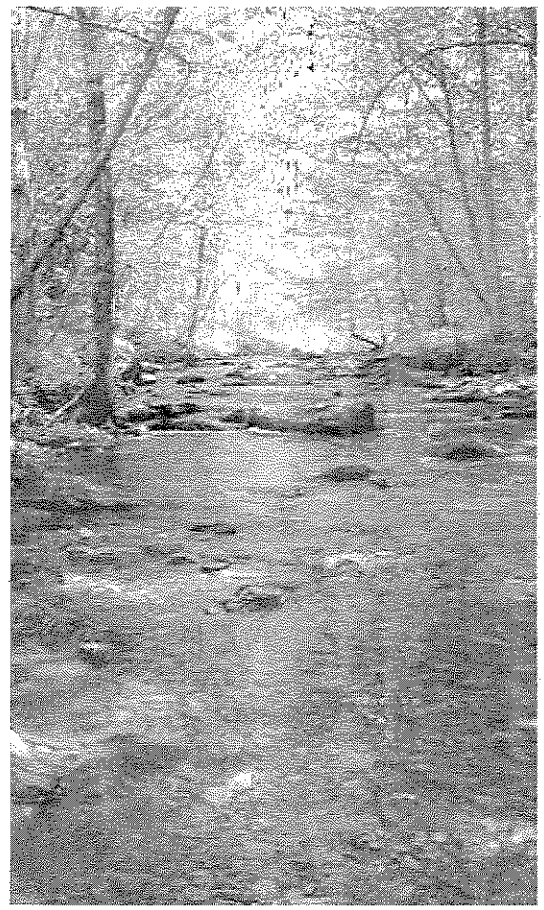
Development of Nutrient Endpoints for Allegheny Plateau and Ridge and Valley Ecoregions of Pennsylvania: TMDL Application

Prepared for

United States Environmental Protection Agency

Region 3

Philadelphia, PA



By

Michael J. Paul and Lei Zheng

Tetra Tech, Inc.

400 Red Brook Boulevard, Suite 200

Owings Mills, MD 21117

June 24, 2008

List of Figures

Figure Page

1 Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts 1

2 Map of the sample sites used in the development of nutrient endpoints using the distribution based approach in this study, labeled by agency affiliation 5

3 Plot of TN and TP samples in the All Sites (All) and two Reference Site (Ref) populations used to estimate candidate endpoints with the distribution based approach 7

4 Response of the three MBSS invertebrate metrics to increases in phosphorus concentration 17

5 Response of three EMAP invertebrate metrics to increases in phosphorus concentration. 18

6 Response of the last EMAP invertebrate metric to an increase in phosphorus concentration 19

List of Tables

Table Page

1 Values of TN and TP candidate endpoints derived using the distribution based approach 6

2 Threshold values for the MBSS and EMAP benthic macroinvertebrate IBI metrics in the Ridge and Valley ecoregion (Southerland et al. 2005, Klemm et al. 2003). 15

3 Summary of candidate endpoints for each of the analytical approaches discussed for the Allegheny Plateau 22

4 Summary of candidate endpoints for each of the analytical approaches discussed for the Ridge and Valley 24

Introduction

The United States Environmental Protection Agency (USEPA) in Region 3 is overseeing the development of nutrient TMDLs to protect aquatic life use for several streams Pennsylvania. Tetra Tech, Inc, (Tt) was approached to establish appropriate and scientifically defensible nutrient endpoints that are protective of aquatic life. Tetra Tech developed endpoints for the Piedmont region as part of this work and published those in a report entitled, "Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania: TMDL Application" dated November 30, 2007. That document described the process that was applied in detail, the results of those analyses, and the final recommended TP and TN endpoints. This addendum applies matching methodology to the development of endpoints for streams of the Ridge and Valley and Allegheny Plateau ecoregions of Pennsylvania.

Nutrients affect aquatic systems in diverse ways, and the effects on most non-primary producer aquatic life uses are indirect (Figure 1).

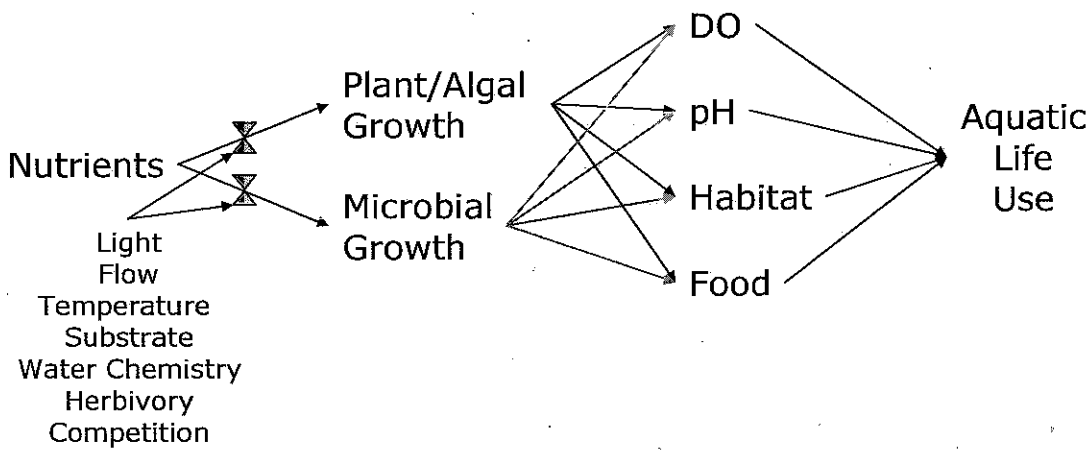


Figure 1 – Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts. Nutrients enrich both plant/algal as well as microbial assemblages, which lead to changes in the physical/chemical habitat and food quality of streams. These effects directly impact the insect and fish assemblages. The effects of nutrients are influenced by a number of other factors as well, such as light, flow, and temperature.

Nutrients cause enrichment of primary producer and decomposer biomass and productivity, the increase of which leads to changes in the physical and chemical stream environment (e.g., reduced oxygen, loss of reproductive habitat, alteration on the availability of palatable algal taxa, etc.). It is these effects which directly result in changes to the biological stream community (e.g., loss of disturbance sensitive taxa), and ultimately impair the use of a stream for aquatic life.

Traditionally, water quality endpoints to protect aquatic life use were developed using toxicological approaches. Such approaches have been applied for a range of pollutants to develop water quality endpoints. However, as explained above, nutrient enrichment does not have a direct toxicological effect on non-primary producer aquatic life. It is worth mentioning that nutrients do, however, affect algal and plant aquatic life directly, altering the diversity and composition of those assemblages radically. For insects, fish and other aquatic life, however, the mode of action of nutrients is indirect and through a causal pathway that involves alteration of physical, chemical, and biological attributes of their habitat. As a result, traditional toxicological approaches are not appropriate.

The USEPA has published guidance on nutrient endpoint development for the protection of designated uses for a range of waterbody types including rivers and streams (USEPA 2000a), but also for lakes and reservoirs (USEPA 2000b), estuaries (USEPA 2001), and wetlands (USEPA 2007). The principal method described in those documents is the use of a frequency distribution-based approach (often called the reference approach), where a percentile of a distribution of values is used to identify a nutrient endpoint. The sample distributions were typically either from least disturbed reference sites (sensu Stoddard et al. 2006) or the entire population of sample sites. These

documents, however, clearly encourage the use of alternative scientifically defensible approaches and, especially, the application of several approaches in a multiple-lines-of-evidence framework, to establish defensible and protective endpoints. The documents state that, "a weight of evidence approach that combines (multiple) approaches…will produce endpoints of greater scientific validity." The approaches recommended include the frequency distribution approach, stressor-response analyses, and literature based values.

In determining nutrient endpoints for developing TMDLs to protect aquatic life uses of Ridge and Valley and Allegheny Plateau streams in Pennsylvania, we relied on a multiple lines of evidence approach framework considering all of the following approaches: frequency distribution based analysis, stressor-responses analyses, and literature based values. The following sections describe these approaches in detail including the methods used for each and the results. The resulting candidate values were then considered and a weight-of-evidence applied to develop final endpoint recommendations.

Due to the limitation of watershed sizes and the difficulty in obtaining stressor response gradients (especially for reference sites) in the target watersheds, we used an ecoregional nutrient endpoint development approach similar to that applied for nutrient criteria development to identify nutrient targets that would protect aquatic life uses in these watersheds. The USEPA, in their recommendations for nutrient endpoint development, specified that "Ecoregional nutrient criteria will be developed to account for the natural variation existing within various parts of the country." (USEPA 2000a)

They go on to explain the importance of ecoregions:

"Ecoregions serve as a framework for evaluating and managing natural resources. The ecoregional classification system developed by Omernik (1987) is based on multiple geographic characteristics (e.g., soils, climate, vegetation, geology, land use) that are believed to cause or reflect the differences in the mosaic of ecosystems."

The two targeted watersheds for this report, Sawmill Run and Paxon Creek, are located within the Allegheny Plateau and Ridge and Valley ecoregions, respectively. We collected data from the same ecoregions in Pennsylvania, Maryland, West Virginia, and Virginia. We made the assumption that nutrient dynamics in the two watersheds should be similar to nutrient dynamics in sites selected from across these two ecoregions, given similarities in geology, soils, and climate.

Frequency Distribution Based Approach

For this approach, we identified water quality samples collected by a variety of agencies from streams in the Allegheny Plateau and Ridge and Valley ecoregions stored in a variety of databases including the USEPA Storage and Retrieval (STORET) and Environmental Monitoring and Assessment Program (EMAP) databases, United States Geological Survey (USGS) National Water Inventory System (NWIS) and National Water Quality Assessment (NAWQA) program, and the Maryland Biological Stream Survey (MBSS) database (Figure 2). Two populations of sites were developed. The first was all sites for which nutrient samples were available (All Sites). The second was all sites for which watershed land cover was available and for which reference criteria could be applied (Reference Sites).

The All Sites population included samples from all of the agencies described above. For sites with multiple samples, samples were averaged to estimate an average site nutrient concentration. This reduced the influence of any one site on the percentiles.

After all the sites were prepared, we calculated the 25th percentile nutrient concentration of total phosphorus (TP) and total nitrogen (TN).

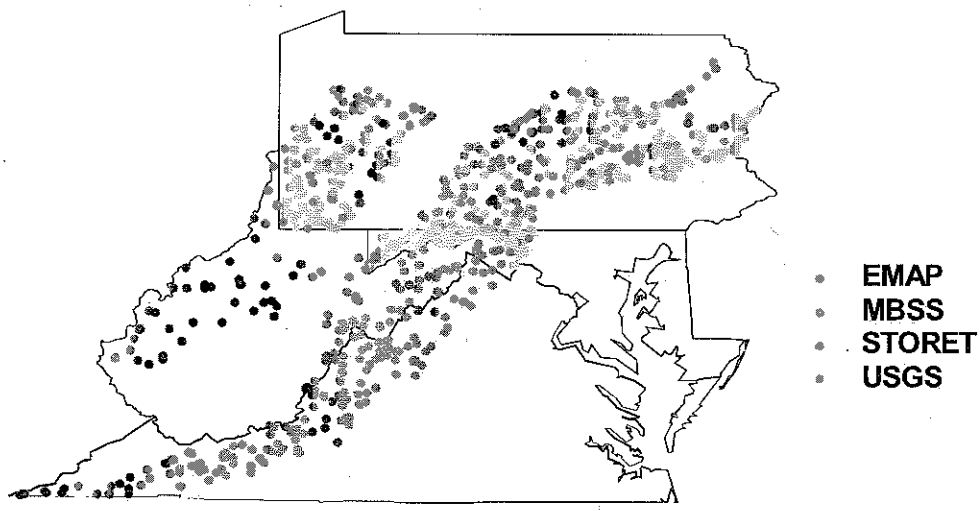


Figure 2 – Map of the sample sites used in the development of nutrient endpoints using the distribution based approach in this study, labeled by agency affiliation.

For sites where land cover information was available (USEPA EMAP, USGS NAWQA, and MBSS), we developed land cover screening criteria to identify least disturbed watersheds (sensu Stoddard et al. 2006). Least disturbed sites represent those watersheds with minimal human disturbance and, therefore, provide the best empirical estimate of chemical integrity. We developed two different reference criteria: >80% Forest, <5% urban (N=7) and >70% Forest, <5% urban (N=24). We then calculated the 75th percentile of total phosphorus and total nitrogen concentrations associated with these populations.

The distribution based analyses resulted in lower endpoints for nutrients from the All Sites population than from the two Reference Site populations in both ecoregions (Figure 3, Table 1). For the Allegheny Plateau ecoregion, total phosphorus endpoints were

between 19 and 36 µg/L and total nitrogen endpoints between 260 and 665 µg/L (Table 1). For the Ridge and Valley ecoregion, distribution based total phosphorus endpoints were between 10 and 15 µg/L and total nitrogen endpoints between 280 and 620 µg/L (Table 1).

Table 1 – Values of TN and TP candidate endpoints derived using the distribution based approach.

	Reference Sites		All Sites
Parameter	>80% Forest <5% Urban	>70% Forest <5% Urban	
	75th Percentile	75th Percentile	25th Percentile
Allegheny Plateau			
TN (µg/L)	425	664	260
TP (µg/L)	36	33	19
N	25	39	125 (TN) 185 (TP)
Ridge and Valley			
TN (µg/L)	480	618	281
TP (µg/L)	13	15	10
N	122	147	885 (TN) 1073 (TP)

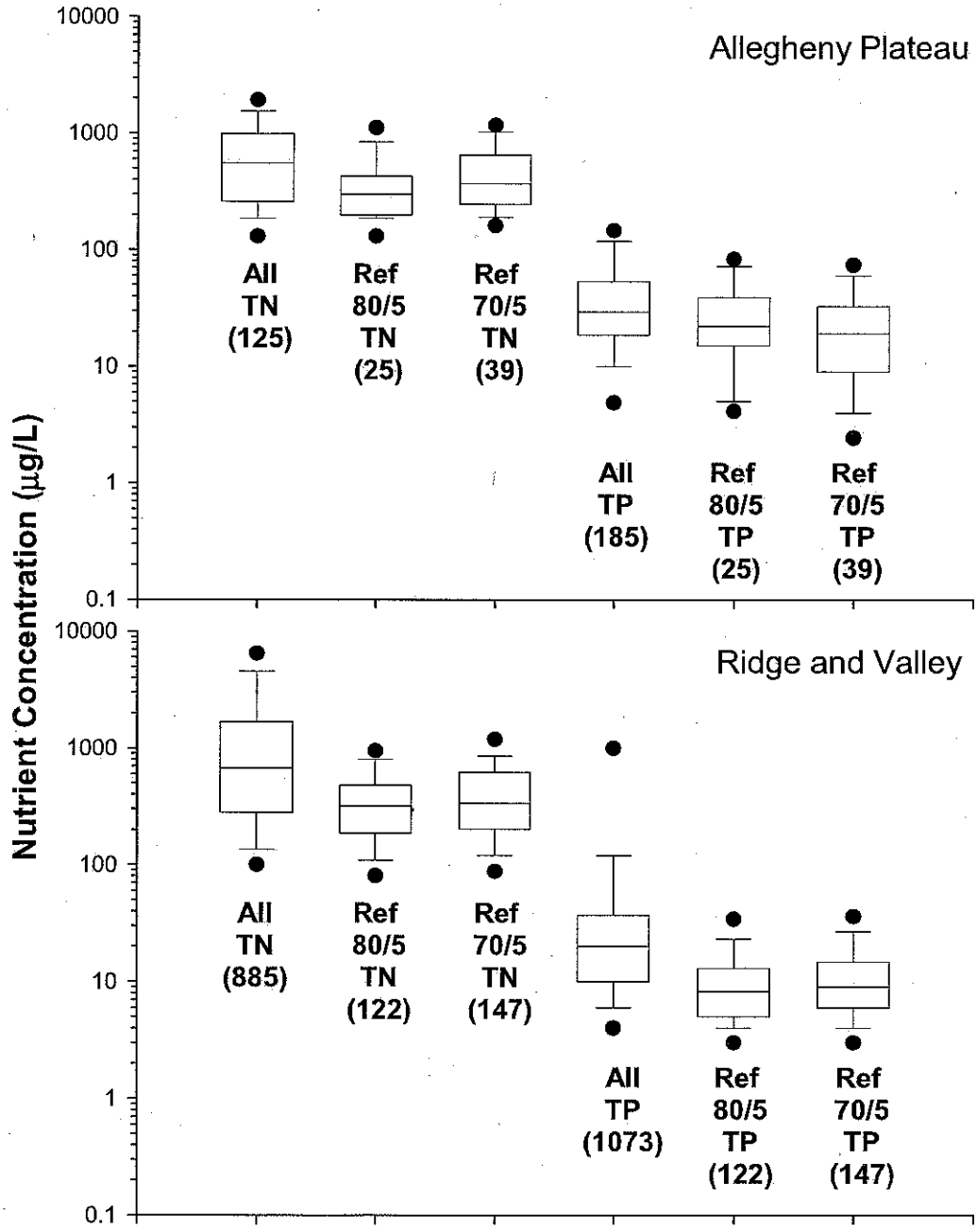


Figure 3 – Plot of TN and TP samples in the All Sites (All) and two Reference Site (Ref) populations used to estimate candidate endpoints with the distribution based approach. Sample sizes are shown below each label. Lines indicate the median values (50th percentiles), boxes are the quartiles (25th and 75th percentiles), whiskers are 10th and 90th percentiles, and symbols are the 5th and 95th percentiles.

Modeled Reference Expectation Approach

Another approach that falls under the rubric of "reference approaches" is the modeled reference expectation approach (Dodds and Oakes 2004). In this approach, multiple regression models of total nutrients versus human land cover (agriculture and urbanization) are built and then solved for the condition of no human land cover (i.e., the intercept). This approach has been used to estimate nutrient concentrations in the absence of human disturbance in the Midwest (Dodds and Oakes 2004).

We developed modeled reference expectation models for the Allegheny Plateau region using data from the USEPA EMAP program, since it was the only one which had both land cover and nutrient data. The final equation for total nitrogen was:

$$\mathrm{Log}_{10}(\mathrm{TN}) = 2.48 + 0.40(\mathrm{arcsine}\sqrt{\%\ \mathrm{Agriculture}}) + 0.94(\mathrm{arcsine}\sqrt{\%\ \mathrm{Urban}});$$

($R^2 = 0.24$; F=9.98, $p<0.001$).

Solving for the undisturbed condition leads to a modeled reference total nitrogen concentration for the Allegheny Plateau of 302 μg/L. No significant model for total phosphorus could be created with the land cover data for the Allegheny Plateau, so we estimated the TP value for this approach using N:P ratios (see below).

Similarly, we developed modeled reference expectation models for TN and TP in the Ridge and Valley ecoregion using data from the USEPA EMAP and Maryland DNR MBSS programs, since they were the only ones which had both land cover and nutrient data. The final equation for TP was:

$$\mathrm{Log}_{10}(\mathrm{TP}) = 0.86 + 0.62(\mathrm{arcsine}\sqrt{\%\ \mathrm{Agriculture}});$$

($R^2 = 0.27$, F=169.0, $p<0.001$)

and for TN was:

$$\text{Log}_{10}(\text{TN}) = 2.32 + 1.01(\text{arcsine}\sqrt{\%\ \text{Agriculture}}) + 0.13(\text{arcsine}\sqrt{\%\ \text{Urban}});$$

($R^2 = 0.51$, F=234.6, $p<0.001$)

Solving for the undisturbed condition leads to a modeled reference Ridge and Valley TP endpoint of 7 µg/L and TN endpoint of 209 µg/L.

N:P Ratios Suggest P Limitation Dominates the Allegheny Plateau Ecoregion

We calculated N:P ratios across all sites in the Allegheny Plateau dataset. The average molar N:P ratio for All Sites was 86:1. We applied this ratio to the TN value estimated from the modeled reference expectation value for TN in the Allegheny Plateau, which yielded a TP value of 8 µg/L TP. The molar ratio of N:P based on the recommended USEPA nutrient criteria for this ecoregion (TP=10 µg/L, TN=310 µg/L) is 68:1. Applying this value, as well as the Redfield molar N:P ratio (16:1), to the value of TN estimated using the modeled reference expectation approach above led to estimated TP values of 10 and 42 µg/L, respectively. We would defend the use of natural ratios rather than Redfield given uncertainties in the applicability of Redfield to freshwater systems combined with the fact that Allegheny Plateau average N:P ratios are much higher than Redfield.

Stressor-Response Approach

Stressor-response approaches refer to a suite of analytical techniques that derive candidate endpoints by exploring the relationships between response variables and nutrient concentrations. Typical response variables in the context of nutrient endpoint development include water chemical aquatic life use indicators (dissolved oxygen, pH, etc.), algal biomass and/or algal assemblage metrics (e.g., percent nutrient sensitive

diatoms), and aquatic life use indicators or biocriteria indicators (e.g., algal multimetric indices or individual metrics scores, invertebrate multimetric indices or individual metrics, etc.). The value of these indicators is their direct linkage to aquatic life use designations. They, therefore, provide a way to connect nutrient concentrations directly to aquatic life use protection. We used a few different stressor-response analytical techniques to develop candidate nutrient endpoints using invertebrate response indicators.

We selected two important nutrient variables to examine biological responses: total nitrogen (TN) and total phosphorus (TP). TN and TP are two of the four primary variables EPA recommended for nutrient endpoint development and are likely to limit aquatic primary producers. TP and TN may reflect stream trophic status better than inorganic P and N because nutrient depletion can be partially offset by increases in particulate fractions of TP and TN resulting from drift and suspension in the water column (Dodds 2002). In addition, TN and TP are also measured more frequently in most of the national and state programs than other nutrient variables.

The primary response variable of interest for stream trophic state characterization is algal biomass, which is most commonly reported as mg/m^2 Chl a. Chl a is a photosynthetic pigment and is a sensitive indicator of algal biomass. It is considered an important biological response variable for nutrient-related problems (USEPA 2000a). Periphyton is also often analyzed for dry mass (DM) and ash free dry mass (AFDM), which includes non-algal organisms. The USEPA also recommends a measure of turbidity as the response variable. However, turbidity is often associated with total suspended solids (TSS) and other environmental factors and is less commonly used as a direct response variable. In addition to these, algal species composition often responds

dramatically to excess nutrients, including the proliferation of eutrophic and nuisance algal taxa. As a result, algal metrics are frequently used as direct indicators of nutrient enrichment (van Dam et al. 1994, Pan et al. 1996). We did not have sufficient algal endpoints to explore these response variables in the Allegheny Plateau and Ridge and Valley ecoregions, as we did for the Piedmont analysis. The aquatic life response variable for which we had sufficient information to consider was macroinvertebrate metrics from multimetric indices. Macroinvertebrate indices are the most reliable and frequently used bioindicators, and many macroinvertebrate metrics are sensitive to nutrient enrichment.

Data:

We collected data from four different national and state programs, similar to those used in the distribution based analyses:

- USEPA Environmental Monitoring and Assessment Program (EMAP)
- USGS National Water Information System (NWIS)
- USEPA STORET database
- Maryland Biological Stream Survey (MBSS) program

Two projects, the USEPA EMAP and MBSS programs, simultaneously collected nutrients and macroinvertebrate composition data, which were valuable for exploring invertebrate assemblage responses to nutrients. The MBSS collected thousands of macroinvertebrate samples from its statewide stream survey including numerous samples in the Ridge and Valley ecoregion and the EMAP Mid-Atlantic Highlands Assessment collected similar samples across both ecoregions throughout Virginia, West Virginia, Maryland, and Pennsylvania.

Data Analysis: Overview

Establishing definitive stressor-response relationships is a valuable line of evidence in the multiple lines of evidence approach. We first used Spearman correlation analysis to examine relationships between response and stressor variables. Correlation analyses identified significant relationships between biological response and nutrient variables. However, correlation may or may not indicate the real relationship. Numerous relationships were examined; only a subset of which were correlated. There were also results that were considered potentially important but showed weaker relationships.

We selected correlations of interest and performed visual scatter plots to further examine the relationships. We used either linear regression or a locally weighted average regression line to examine the trend of change along the environmental gradients. The locally weighted scatterplot smoothing (LOWESS) technique (Cleveland 1979) models nonlinear relationships where linear methods do not perform well. LOWESS fits simple models to localized subsets of the data to construct a function that describes, essentially, the central tendency of the data. LOWESS fits segments of the data to the model. Tension, which describes the portion of data being used to fit each local function, was set at 0.50 for LOWESS regression.

We also used conditional probability analysis (Paul and MacDonald, 2005) to examine changes in the biological community along stressor gradients. Conditional probability provides the likelihood (probability) of a predefined response when a specific value of a pollutant stressor (condition) is exceeded. Conditional probability is the likelihood of an event when it is known that some other event has occurred. Conditional probability answers the question: for a given threshold of a stressor, what is the cumulative probability of impairment? For example, if the total phosphorus

concentration is greater than 30 μg/L, what is the probability of biological impairment (defined as < 8 EPT Taxa) for each site under consideration? All observed stressor values (in this example, all observed values of total phosphorous) are used to develop a curve of conditional probability (Paul and MacDonald, 2005). Because of its ability to identify risks of impact associated with given nutrient concentrations, the approach is suited to identifying nutrient thresholds protective of aquatic biological condition.

To estimate conditional probability of an impairment, we first had to define impairment as a specific value for a response variable (e.g., EPT < 8 genera). We used preexisting biocriteria thresholds as our response thresholds (MDNR 2005). For the Ridge and Valley ecoregion, we used MBSS and EMAP data as well as criteria based on scoring thresholds developed by the state of Maryland for their multimetric index and by EPA EMAP for use in their multimetric index for the Mid-Atlantic Highlands (Klemm et al. 2003). For the Allegheny Plateau, we used EMAP metrics alone because MBSS did not sample in this region. Thresholds used for the EMAP metrics were the 25th percentile of reference site metric scores for metrics declining with stress and the 75th percentile of reference sites for metrics increasing with stress. These thresholds are commonly used to identify metrics that discriminate between reference and stressed sites (Barbour et al. 1999). We used the same reference criteria developed by Klemm et al. (2003) except we excluded the nutrient criteria they used (to avoid circularity) and used only their cutoffs for chloride, sulfate, acid neutralizing capacity, and habitat.

We also used nonparametric deviance reduction (change point analysis) to identify thresholds in biological responses to nutrients (Qian et al. 2003). This technique is similar to regression tree models, which are used to generate predictive models of response

variables for one or more predictors. The change-point, in our application, is the first split of a tree model with a single predictor variable (nutrient concentration). The loss function of regression trees can be evaluated by the proportion of reduction in error (PRE), which is analagous to the multiple R^2 of general linear models.

Data Analysis: Metric Calculation

Macroinvertebrate Metrics

Numerous macroinvertebrate assemblage metrics were assembled from the MBSS and EMAP programs. We selected a subset of benthic macroinvertebrate indicators, focusing on those that composed the MBSS IBI (Ridge and Valley) and/or Highlands EMAP IBI (Allegheny Plateau). Metrics considered included Ephemeroptera, Plecoptera, Trichoptera (EPT) richness, Ephemeroptera Richness, Plecoptera Richness, Trichoptera Richness, Tolerant Richness, Percent Tolerant, Scraper Richness, Percent Scrapers, Collector-Filterer Richness, and Percent Dominant 5 taxa.

Results: Benthic Macroinvertebrate Metrics – Nutrient Relationships

The largest datasets available for analyzing macroinvertebrate responses to nutrient concentrations were the Maryland Biological Stream Survey (MBSS) and EMAP Mid-Atlantic Highlands Assessment datasets. We found 50 samples from the EMAP database with corresponding macroinvertebrate metric and nutrient data for the Allegheny Plateau ecoregion. In contrast, we found 242 samples with corresponding macroinvertebrate metrics and nutrient samples from the MBSS dataset, and 320 comparable samples from the EMAP database in the Ridge and Valley ecoregion. For each metric, scoring criteria were developed based on the distribution of values from least disturbed reference sites (Table 2). For the MBSS, we selected the middle point of the distribution as the

impairment threshold for each metric, since this is consistent with their methodology (Southerland et al. 2005, Table 2). For the EMAP data, we used a standard practice, namely using the 25th percentile of reference site metric scores (for metrics decreasing with stress) or the 75th percentile of reference site metrics cores (for metrics increasing with stress) as our thresholds (Table 2, Barbour et al. 1999).

Of the metrics considered in the Allegheny Plateau, none exhibited a strong enough response to nutrient concentrations to merit development of potential endpoints using the stressor-response approach. For the Ridge and Valley, however, several exhibited a strong response to TP and we used the following metrics: MBSS – EPT Richness, Percent Scrapers, and Number of Taxa; EMAP – EPT Richness, Ephemeroptera Richness, Trichoptera Richness, and Percent Dominant 5 Taxa.

Table 2 – Threshold values for the MBSS and EMAP benthic macroinvertebrate IBI metrics in the Ridge and Valley ecoregion (Southerland et al. 2005, Klemm et al. 2003).

MBSS Scoring criteria	5	3	1	Mid Point
Number of Taxa	≥ 24	15 – 23	<15	19
Number of EPT	≥ 14	8 – 13	< 8	10.5
% Scrapers	≥ 13	3 – 12	< 3	7.5
EMAP Scoring criteria	25th Percentile of Reference		75th Percentile of Reference	
Number of EPT	16			
Number of Ephemeroptera	7			
Number of Trichoptera	4			
Percent Dominant 5 Taxa			60.75	

MBSS Metrics

The three MBSS metrics (Total Taxa, EPT Taxa, and Percent Scrapers) all declined with increased TP concentrations (Figure 4). The scatterplots exhibited a traditional wedge shape decline, while the conditional probability graphs clearly indicated the probability of impairment increasing as TP concentrations increased from 10 to 50 μg/L

TP. Change point analyses indicated thresholds at 14, 14, and 16 µg/L TP for these three metrics, respectively.

EMAP Metrics

Similarly, the first three EMAP metrics all declined with increasing TP concentrations, also exhibiting the typical wedge shaped response (Figure 5). The same data expressed as conditional probabilities exhibited increasing risk of impacts between 8 and 50 µg/L TP. Change point analyses indicated thresholds at 19 µg/L TP for all three metrics. The last EMAP metric, Percent Dominant 5 taxa, increased with increasing TP concentration, as expected (Figure 6). As macroinvertebrate communities become stressed, there is a predictable decline in diversity and evenness, as a few tolerant taxa (e.g., weedy species), take advantage of the loss of more sensitive taxa and begin to dominate the assemblage (Klemm et al. 2003). Change point analyses indicated a threshold at 23 µg/L TP for this metric.

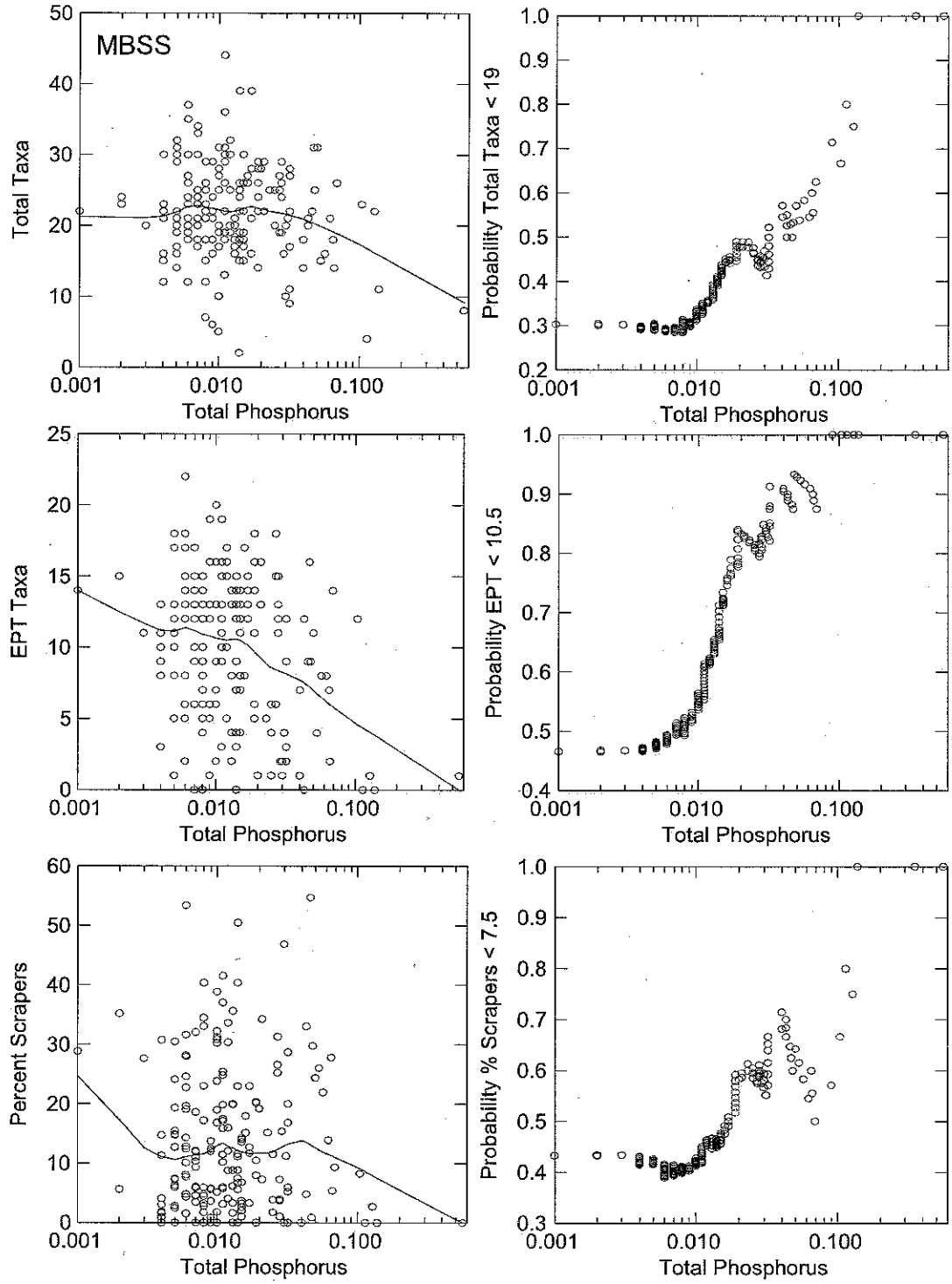


Figure 4 – Response of the three MBSS invertebrate metrics to increases in phosphorus concentration. Plots on the left are raw data with a lowess curve fit. Plots on the right are the same raw data expressed as conditional probabilities.

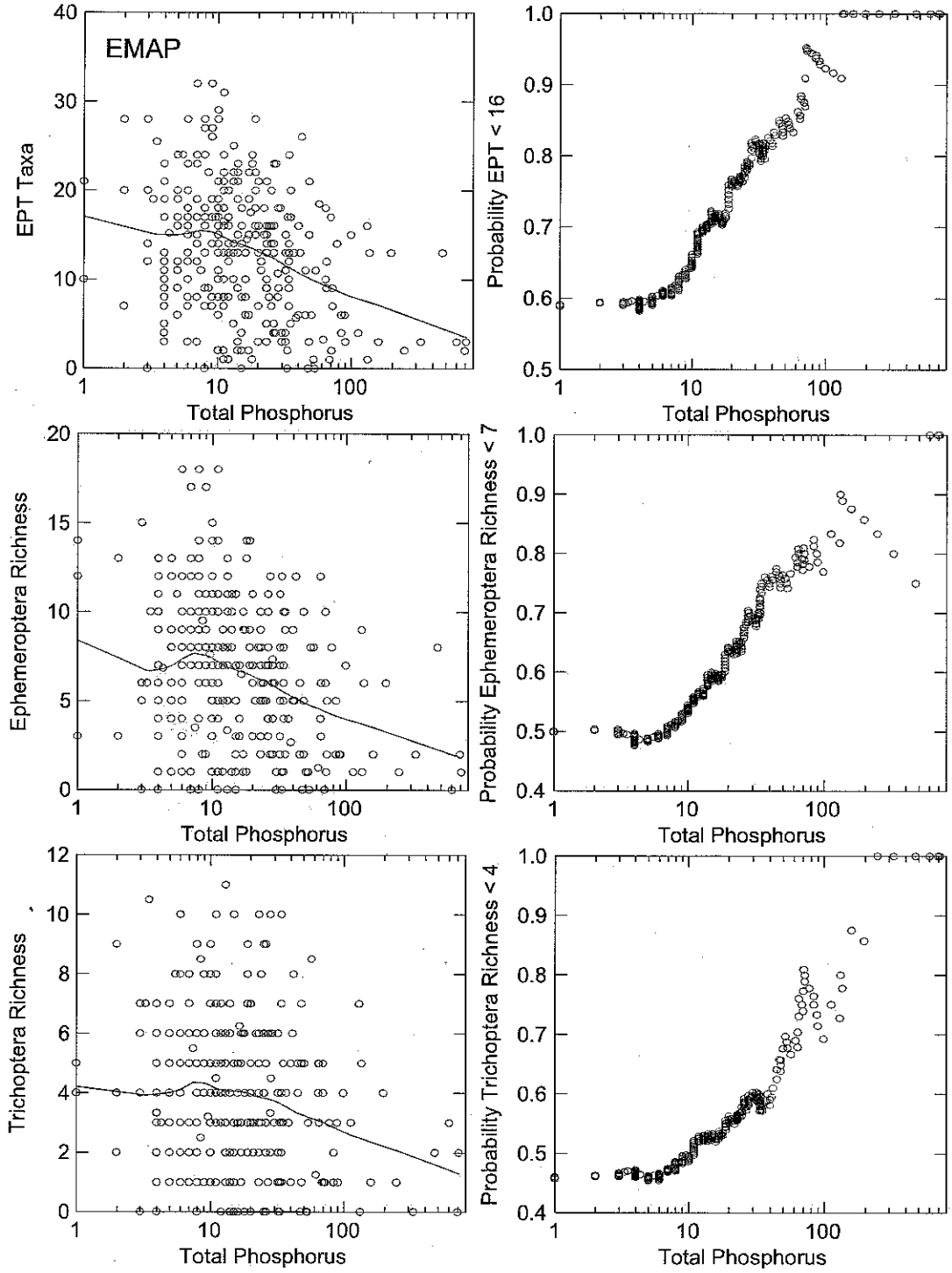


Figure 5 - Response of three EMAP invertebrate metrics to increases in phosphorus concentration. Plots on the left are raw data with a lowess curve fit. Plots on the right are the same raw data expressed as conditional probabilities.

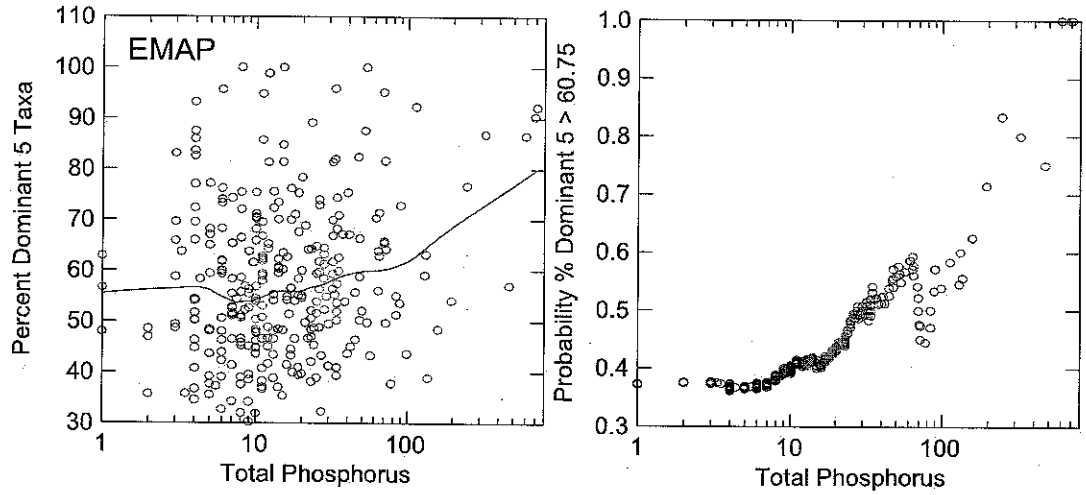


Figure 6 - Response of the last EMAP invertebrate metric to an increase in phosphorus concentration. Plot on the left is raw data with a lowess curve fit. Plot on the right is the same raw data expressed as conditional probabilities.

Literature Based Analysis: Current Existing Endpoints or Threshold Values

In this last analytical section, we present several studies relevant to the development of nutrient endpoints in the Allegheny Plateau and Ridge and Valley ecoregions of Pennsylvania. These are taken principally from the peer-reviewed and federal agency technical literature and reflect increasing experimental and theoretical interest in the impact of nutrients on natural stream systems. We attempted to extract information from these studies that could recommend specific endpoints.

In natural, shaded streams [such as those evaluated in the Dodds et al. (2002) model], it is difficult to assess the full growth potential of algae. Algal growth potential has been evaluated using artificial stream channels that are fully exposed to nutrient and light gradients. Previous studies (Horner et al. 1983, Bothwell 1989) demonstrated that in artificial streams, algal growth could be saturated (i.e., achieved maximum growth rate) at 25–50 µg/l phosphorus. Rier and Stevenson (2006) found that at 16 µg/L soluble

reactive phosphorus (SRP) or 86 µg/L dissolved inorganic nitrogen (DIN), algal growth was at 90% of its maximum rate. They also found that saturation concentrations were 3–5 times lower than concentrations needed to produce maximum algal biomass (i.e., 430 µg/L DIN and 80 µg/L SRP for growth saturation). However, these values were derived mostly on the basis of diatom and bluegreen algae growth. We expect that green algae (i.e., *Cladophora*) would have higher nutrient saturation concentrations for peak growth (Borchardt 1996).

USEPA's nutrient threshold recommendations for the Allegheny Plateau and Ridge and Valley nutrient ecoregion were 310 µg/L for TN and 10 µg/L for TP.

Dodds and Welch (2000) conducted a meta-study including values from a range of areas nationwide. These were combined into regression equations to predict chlorophyll. They found that if a mean of 50 mg/m^2 of chlorophyll is the target (thus insuring chlorophyll is less than 100 mg/m^2 most of the time), TN should be 470 µg/L and TP should be 60 µg/L. Even lower numbers should be considered for more pristine waters. These estimates were more general in scope. These authors further noted that lower TN and TP values associated with these chlorophyll concentrations were obtained when using a detailed, smaller data set than those from a larger data set (55 µg/L TP from a large dataset versus 21 µg/L for a more specific, local data set).

USGS conducted a study in 2001 for a broad area of the US, including the New River and Big Sandy River in Virginia (Robertson et al. 2001). They looked at 234 sites using the reference approach and found that a TP of 20 µg/L was appropriate for what they define as Environmental Nutrient Zone 2.

Rohm et al. (2002) conducted a national study to demonstrate how regional reference conditions and draft nutrient endpoints could be developed. They divided the country into 14 regions and analyzed available nutrient data as a case study, using EMAP data from Central and Eastern Forested Uplands, an area that includes much of central Pennsylvania. This case study suggested a criterion of 375 µg/L for TN and 13 µg/L for TP. Rough estimates from the data presented for their Region IX that includes Eastern Pennsylvania gives estimates of 500 µg/L TN and 20 µg/L TP.

Recommended Endpoints

Total Phosphorus (TP)

Endpoint (magnitude) – Allegheny Plateau

Our analyses relied on a weight-of-evidence analysis drawing on many different analytical approaches. Each of the different approaches produced slightly different endpoints and these are summarized in Table 3.

In a weight-of-evidence approach, the different analyses are weighted on their applicability and the strength of the analysis. For the Allegheny Plateau, we had insufficient data to produce significant stressor-response relationships. As a result, we were left weighting the distribution based, modeled reference expectation, and scientific literature lines.

For the distribution based approach, we assembled a large population of nutrient concentration from sites ranging in quality from various databases. We identified the entire population of sites for one estimate, and identified a subset of minimally disturbed

sites for a second estimate. The values estimated from these populations were between 19 and 36 µg/L TP.

The modeled reference expectation did not produce a significant model for TP in this ecoregion, but did for TN (302 µg/L). We used the molar ratio of N:P to identify an appropriate TP target associated with this TN concentration. The average Allegheny Plateau stream N:P from our dataset was 86:1. The ratio of N:P based on USEPA's recommended endpoints was similar (68:1). Using these two, and the Redfield ratio (16:1), resulted in TP endpoints of 8, 10, and 42 µg/L respectively.

Finally, literature relevant to nutrient endpoints for this region ranged from approximately 10 µg/L TP (USEPA recommended criteria) to 60 µg/L (Dodds and Welch 2000), but most values were centered around 30 µg/L TP.

We weighted the reference criteria line of evidence most highly of the three lines we had available and we recommend a TP endpoint of 35 µg/L TP for streams of this region.

• Allegheny Plateau recommended endpoint: 35 µg/L TP

Table 3 – Summary of candidate endpoints for each of the analytical approaches discussed for the Allegheny Plateau.

	Approach	TP Endpoint (µg/L)
Reference Approach		**19-36**
	Reference Site 75th Percentile	33-36
	All Sites 25th Percentile	19
Modeled Reference		**8-42**
Stressor-Response		**NA**
Other Literature		**13-100**
	USEPA Recommended Regional Criteria	10

USEPA Regional Criteria Approach – Local Data	13
Algal Growth Saturation	25-50
Nationwide Meta-Study TP-Chlorophyll	21-60
USGS Regional Reference Study	20
USGS National Nutrient Criteria Study	13-20

Endpoint (magnitude) – Ridge and Valley

As above, our analyses relied on a weight-of-evidence analysis drawing on many different analytical approaches. Each of the different approaches produced slightly different endpoints and these are summarized in Table 4.

In a weight-of-evidence approach, the different analyses are weighted based, essentially, on their applicability and the strength of the analysis. For the Ridge and Valley, we had substantially more data, including abundant data on stressor-response relationships. As a result, we were able to use all four lines of evidence: distribution based, modeled reference expectation, stressor-response and scientific literature based approaches.

Similar to the Allegheny Plateau distribution based approach, we assembled a large population of nutrient concentrations from various databases for sites ranging in quality. We identified the entire population of sites for one estimate, and identified a subset of minimally disturbed sites for a second estimate. The values estimated from these populations were between 10 and 15 µg/L TP in this ecoregion.

The modeled reference expectation produced significant models for both TP and TN in this ecoregion, so we did not have to rely on N:P ratios to estimate a TP endpoint using this line of evidence. The TP endpoint from modeled reference expectation was 7 µg/L. The TN generated from this approach was 209 µg/L. Most streams in the Ridge and Valley ecoregion, similar to the Allegheny Plateau and Piedmont, appear to be P limited

systems. The median N:P ratio across the streams sampled was well above Redfield (16:1) and was actually 88:1. Using this ratio along with the TN endpoint, the TP endpoint would be 5 µg/L. Using the more conservative Redfield Ratio (16:1), combined with the TN endpoint, results in a TP value of 29 µg/L.

The stressor-response analyses led to a variety of endpoints that varied between 14 and 23 µg/L TP. The lowest threshold (14 µg/L) was observed in the EPT taxa response for the MBSS data and the highest threshold for the Percent Dominant 5 Taxa metric from the EMAP dataset (23 µg/L).

Finally, literature relevant to nutrient endpoints for this region ranged from approximately 10 µg/L TP (USEPA recommended criteria) to 60 µg/L (Dodds and Welch 2000), but most values were centered around 30 µg/L TP.

We weighted the stressor-response line of evidence most highly of the four lines we had available as it provided a direct linkage to use measures, and these results were higher than the distribution based and modeled reference values. Literature based values were, in terms of central tendency, closer to the upper end of the stressor-response derived values. As a result, we recommend a TP endpoint of 25 µg/L TP for streams of the Ridge and Valley.

• Ridge and Valley recommended endpoint: 25 µg/L TP

Table 3 – Summary of candidate endpoints for each of the analytical approaches discussed for the Ridge and Valley.

	Approach	TP Endpoint (µg/L)
Distribution Based		**10-15**
	Reference Site 75th Percentile	13-15
	All Sites 25th Percentile	10

Modeled Reference		**10-15**
Stressor-Response		**14-23**
	MBSS	
	Total Taxa	14
	EPT Taxa	14
	Percent Scrapers	16
	EMAP	
	EPT Taxa	19
	Ephemeroptera Taxa	19
	Trichoptera Taxa	19
	Percent Dominant 5 Taxa	23
Other Literature		**13-100**
	USEPA Recommended Regional Criteria	10
	USEPA Regional Criteria Approach – Local Data	13
	Algal Growth Saturation	25-50
	Nationwide Meta-Study TP-Chlorophyll	21-60
	USGS Regional Reference Study	20
	USGS National Nutrient Criteria Study	13-20

Sample period

We recommend applying the endpoint over the algal growing season (April to October), which in streams is typically the time during which the greatest risk of deleterious algal growth exists.

• Endpoint applies from April to October

Sample duration

Unlike toxics, there is less literature to recommend appropriate sample duration and frequencies for nutrients. Toxics, with chronic and acute criteria, have a longer history of implementation. Their mode of action is also very different than nutrients. As a result, it was more difficult to recommend an appropriate sample period than to derive the endpoints themselves.

Humans tend to sample nutrients at temporal scales that are different than those to which stream organisms respond. Streams respond both to pulsed as well as chronic nutrient concentrations. For example, algae possess mechanisms to store nutrients and use these stored nutrients for growth over time – so they can respond to episodic inputs. Moreover, the responses to episodic inputs include both assemblage responses (for example, development the nuisance algal taxa) as well as population and individual responses (biomass).

The nutrient data we analyzed for the invertebrate and plant responses were based primarily on single grab samples associated with biological sampling. These analyses, therefore, represent a space for time substitution of sorts, estimating what would occur in a piedmont stream as nutrient concentrations increase.

These factors would recommend a not-to-exceed criterion. However, water velocity affects nutrient delivery in streams and elevated nutrients associated with high flows may not be as accessible to benthic algae. We also recognize that there is resistance to not-to-exceed standards and concern about the risk of capturing false positives, even though the risk of false negatives is similarly great. These concerns would recommend averaging multiple samples over some time period. Algal and microbial responses to nutrients can occur rapidly, but these can be offset by floods that scour the bottom and remove algae. At this time, there is limited information and we have had insufficient time to investigate appropriate averaging periods, especially those that result in conditions detrimental to uses.

As a result, for the purposes of these TMDLs, we recommend that the TP

• Endpoint is assessed as the average TP concentration during the growing period over one year.

endpoint be applied as an average of water samples taken over the growing season. Realize, again, that there is less information to guide this recommendation, which is based principally on our professional judgment and in an attempt to be consistent with other typical duration procedures. A more conservative alternative would be to use the recommended endpoint as a not-to-exceed value, but again, we have had insufficient time to evaluate this.

We feel that this approach will be protective, but we strongly encourage the state and USEPA to investigate this issue more fully for the purposes of regional criteria development. For the TMDLs, this approach is sufficient, but it deserves more attention and resources before being applied to regional criteria.

Literature Cited

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Revision to rapid bioassessment protocols for use in stream and rivers: periphyton, BMIs and fish. EPA 841-D-97-002. U.S. Environmental Protection Agency. Washington DC.

Borchardt, M. A. 1996. Nutrients. Pages 183–227 in R. J. Stevenson, M. L. Bothwell, and R. L. Lowe, editors. Algal ecology: freshwater benthic ecosystems. Academic Press, San Diego.

Bothwell, M. L. 1989. Phosphorus Limited Growth Dynamics of Lotic Periphytic Diatom Communities - Areal Biomass and Cellular Growth-Rate Responses. Canadian Journal of Fisheries and Aquatic Sciences 8(46): 1293-1301.

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

Dodds, W.K. 2002. Freshwater Ecology: Concepts and Environmental Applications. Academic Press, San Diego.

Dodds, W. K. and R. M. Oakes,. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography-Methods (2): 333-341.

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

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

Dodds, W.K. and E.B. Welch. 2000. Establishing nutrient criteria in streams. Journal of the North American Benthological Society 19:186-196

ENSR, International. 2003. Collection and evaluation of ambient nutrient data for rivers and streams in New England. New England Interstate Water Pollution Control Commission, Technical Report. http://www.neiwpcc.org/ncreports.asp

Hill, J. and G. Devlin. 2003. Memorandum: Associations Between Biological, Habitat, and Ambient Data in Upland Western Virginia Ecoregions. Virginia Department of Environmental Quality, West Central Regional Office. Roanoke, VA. (December 17, 2003).

Horner, R.R., E.B. Welch, and R.B. Veenstra. 1983. Development of nuisance periphytic algae in laboratory streams in relation to enrichment and velocity. Pages 121-134 in

R.G. Wetzel (ed.). Periphyton of Freshwater Ecosystems. Dr. W. Junk Publishers. The Hague.

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

Maryland Department of Natural Resources (MDNR). 2005. New Biological Indicators to Better Assess the Condition of Maryland Streams. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment, Annapolis, MD. CBWP-MANTA-EQ-05-13.

Pan, Y. D., R. J. Stevenson, B. H. Hill, A. T. Herlihy and G. B. Collins,. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. Journal of the North American Benthological Society 4(15): 481-495.

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

Ponader, K. & D. F. Charles. 2003. Understanding the relationship between natural conditions and loadings on eutrophication: Algal indicators of eutrophication for New Jersey streams. Final Report Year 2. Report No. 03-04. Patrick Center for Environmental Research, Academy of Natural Sciences. Philadelphia, PA.

Ponader, K., C. Flinders, and D. Charles. 2005. The Development of Algae-based Water Quality Monitoring Tools for Virginia Streams. Report No. 05-09 for the West Central Regional Office, Virginia Department of Environmental Quality. Patrick Center for Environmental Research, Academy of Natural Sciences. Philadelphia, PA.

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

Redfield, A.C. 1958. The biological control of chemical factors in the environment. American Scientist 46, 205-221.

Rier, S. T. and R. J. Stevenson. 2006. Response of periphytic algae to gradients in nitrogen and phosphorus in streamside mesocosms. Hydrobiologia 561:131-147.

Robertson, D.M., D.A. Saad, A.M. Wieben, U.S. Environmental Protection Agency Region V, U.S. Environmental Protection Agency Region VII, and U.S. Geological Survey. 2001. An Alternative Regionalization Scheme for Defining Nutrient Criteria

for Rivers and Streams. U.S. Geological Survey Water-Resources Investigations Report 01-4073. Middleton, WI.

Rohm, C.M., J.M. Omernik, A.J. Woods, and J.L. Stoddard. 2002. Regional characteristics of nutrient concentrations in streams and their application to nutrient criteria development. Journal of the American Water Resources Association 38: 213-239.

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

United States Environmental Protection Agency (USEPA). 2000a. Nutrient Criteria Technical Guidance Manual. Rivers and Streams. United States Environmental Protection Agency, Office of Water, Washington, DC. EPA-822-B-00-002

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

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

United States Environmental Protection Agency (USEPA). 2007. Nutrient Criteria Technical Guidance Manual. Estuarine and Coastal Marine Waters. United States Environmental Protection Agency, Office of Water, Washington, DC. EPA-822-R-07-004

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

HALL & ASSOCIATES

Suite 203
1101 15th Street, N.W.
Washington, D.C. 20005-5004
Telephone: 202-463-1166 Web Site: http://www.hall-associates.com Fax: 202-463-4207
Reply to E-mail:

jhall@hall-associates.com

August 21, 2008

Via Email and U.S. Mail
Administrator Stephen L. Johnson
Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue NW
Washington, DC 20460

RE: Request for Peer Review of New EPA Region III Approach to Developing Instream Standards for Nutrients

Dear Administrator Johnson:

On behalf of the Communities located in three Pennsylvania watersheds (Chester, Paxton and Indian Creeks) and in accordance with the OMB's document entitled *Final Information Quality Bulletin for Peer Review*, we are requesting that EPA initiate a formal independent peer review of the unprecedented scientific approach EPA Region III utilized to develop stringent phosphorus stream standards for three major Pennsylvania ecoregions. The new standards will result in billions of dollars in additional treatment costs in Pennsylvania. Moreover, the standards are physically unattainable by the MS4 communities and non-point sources. As compliance with these requirements is impossible to achieve, severe economic impacts are expected to occur if these objectives are enforced. Consequently, it is imperative that the technical validity for this new approach be independently assessed.

The Region's unprecedented approach for developing stream nutrient standards does not consider whether nutrient control will limit plant growth, contrary to all published EPA Section 304(a) criteria development guidance for nutrients. The new approach *assumes* phosphorus directly impairs sensitive invertebrate populations regardless of its impact on aquatic plants, a concept previously rejected in EPA's nutrient criteria guidance. Moreover, the new statistical method (conditional probability) employed to generate the instream standard neither (1) confirms nutrients are the cause of any impairment nor (2) demonstrates that the Region's new use impairment of concern (invertebrate diversity) will be remedied by nutrient control. In short, the new procedures are a radical departure from published criteria development methods and historical approaches used to generate

Section 303(c) and 304(a) criteria that have always been premised on a clear scientific demonstration of causation and need.

The new technical procedures employed by the Region have never been documented or peer reviewed by EPA to be a sufficient or scientifically defensible basis to generate a necessary and protective Section 304 (a) numeric stream criterion. Nonetheless, the Office of Water informally determined that the new approach was a proper basis for developing Section 304(a) criteria and for imposing expansive point and non-point control measures for phosphorus reduction. (Attachment, Exhibit 8) Once our group realized the Office of Water was intent on supporting the Region's unprecedented approach, we contacted two internationally recognized experts, Dr. Dominic DiToro and Dr. Stephen Chapra, who EPA has often relied upon regarding nutrient issues, to ask for their opinion on the new approach. Their opinions were offered without request for compensation and were made solely due to their professional interest in ensuring scientifically defensible and cost effective approaches are used in nutrient regulation. Both experts clearly and very adamantly concluded the new approach and resulting numeric water quality standards (1) are not scientifically defensible, (2) would not likely restore stream impairments, and (3) would misdirect local resources. (Attachment, Exhibits 9 and 10) In particular, Dr. Chapra concluded that "This is such a scientifically indefensible representation of the connection between nutrients and ecosystem health that I believe that its adoption would represent a grave mistake.... I am much more concerned that its adoption would ultimately be ineffective. That is, it could lead to costly controls that would not protect our precious stream ecosystems." Both scientists stated that either a Science Advisory Board or National Academy of Sciences independent peer review of the new approach should occur given the new procedure's radical departure from accepted methods for criteria derivation and nutrient impact assessment.

Federal Policy and Guidance Mandate Peer Review

As you are aware, EPA has historically conducted an independent peer review of new scientific approaches before utilizing such approaches in the water quality criteria development process. (See, e.g., Science Advisory Board review of EPA's Approach to Emerging Contaminants and EPA's 2006 Peer Review Handbook). As stated in EPA's Handbook, "The principle underlying the Peer Review Policy is that all influential scientific and technical work products used in decision making will be peer reviewed." (Handbook @ 30) The purpose of the independent peer review is, at a minimum, to ensure EPA is basing its regulatory program requirements on scientifically-defensible and well-documented evidence linking the environmental concern to a workable regulatory solution. Virtually every criterion identified in EPA's Handbook (@ 30-32), to determine if a peer review is necessary, is met in this case: the technical work product uses a new scientific approach establishing a precedent that departs significantly from prior agency approaches and has widespread regulatory implications and cost impacts.

Moreover, OMB has adopted mandatory review procedures that are applicable in this instance given the controversy, high cost, and precedent setting nature of the Region III and Office of Water actions. (See, OMB Bulletin – *Final Information Quality Bulletin for Peer Review,* January 14, 2005). EPA plainly intends for the new procedures to be used on a nationwide basis for deriving stream nutrient standards, as acknowledged in the Office of Water memorandum. We understand that states have been requested to use this new approach in developing nutrient standards; therefore, such procedures constitute "highly influential scientific assessments" subject to the strictest peer review requirements. This new standards derivation procedure should have been peer reviewed *before* it was applied to Pennsylvania communities. Nonetheless, it is still subject to the OMB peer review requirement and that action should be promptly undertaken. (See OMB Bulletin, 70 Fed. Reg. 2673-2674, January 14, 2005)

The nationwide application of these new methodologies will result in stream nutrient standards that are beyond the limits of technology and are unattainable in most instances. The cost of compliance will easily exceed several hundred billion dollars. Given the nationwide importance of having scientifically-defensible procedures for generating stream nutrient standards, we respectfully request that you direct the Office of Water to submit the new approach for independent peer review at the National Academy of Sciences. Moreover, as it is highly probable that the new, more restrictive phosphorus reduction requirements mandated by EPA Region III's recent TMDL actions will not result in any meaningful ecological improvements, we ask that you stay application of those TMDLs in the NPDES process until the required peer review is completed.

The attached report, "Justification for Peer Review of EPA's New Methodology for Nutrient Criteria Development", provides additional background information on the history of the Region's TMDL development in Pennsylvania. Thank you for your consideration of this request.

Sincerely,

John C. Hall

Attachments

cc: Senator Arlen Specter
Senator Robert P. Casey
Benjamin Grumbles, USEPA
Pennsylvania Periphyton Coalition
Pennsylvania Municipal Authorities Association
Pennsylvania Water Environment Association
Home Builders Association of Metropolitan Harrisonburg

JUSTIFICATION FOR PEER REVIEW OF EPA'S NEW METHODOLOGY FOR NUTRIENT CRITERIA DEVELOPMENT

Background on EPA Region III Nutrient Criteria Development in Pennsylvania

Since 2005, EPA Region III has encountered repeated difficulties in developing a scientifically defensible approach to determine the effects of nutrients on plant growth in streams. These difficulties, as discussed below, led the Region to abandon accepted methodologies for determining whether nutrients were causing stream use impairment (i.e., excessive plant growth). The new approach assumes that nutrients have a direct effect on invertebrate populations, as if this constituent was a toxic substance. A statistical method known as conditional probability was used to assess field data wherein it was assumed that nutrients were the direct cause of any monitored changes in invertebrate populations. Using the results of this analysis in a "weight-of-evidence" approach (that included three other methods that were divorced from any showing that nutrient levels caused changes in invertebrate populations), the Region selected a "growing season" total phosphorus level and asserted it would ensure the impaired invertebrate populations in the stream would be restored. The available site-specific data on invertebrate populations and nutrient levels were not used to assess whether or not the new approach produced a rational, scientifically defensible result.

The following evaluation reviews applicable regulatory requirements for criteria derivation and then details the fundamental changes the Region and EPA Headquarters made to those procedures in their effort to justify nutrient reduction for stream discharges. These fundamental changes were undertaken using new scientific methodologies that have never undergone any independent scientific peer review to determine their sufficiency for generating appropriate and protective stream nutrient standards.

Statutory and Regulatory Requirements for Criteria Derivation

By statute, criteria must be based on the latest available science and set at the level necessary to protect aquatic life and human health uses. CWA Section 304(a). To achieve this requirement it is essential that criteria possess two attributes: (1) the criteria must be based on data that confirm the pollutant is causing use impairment at ambient concentrations, and (2) the level at which the numeric criteria is set is both sufficient and necessary to protect stream uses. Thus, criteria are, in general, set at the threshold level where the pollutant exposure is demonstrated not to pose a significant threat to aquatic life. (Section 304(a); 40 CFR 131.2 131.3 (b), (c))

Since 1985, EPA has had a well defined procedure for developing water quality criteria when it published the document entitled "*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses,* " USEPA 1985 (hereafter "*Guidelines*"). The *Guidelines* establish a number of very

specific scientific thresholds that must be met to establish criteria that meet Section 304(a) mandates, as follows:

- Water quality criteria must ensure use protection "with a small probability of considerable overprotection or under-protection." (*Guidelines* @ 5)

- It is not enough that the criterion is the best estimate given the available data. Criteria should be derived "only if adequate appropriate data are available to provide reasonable confidence that it is a good estimate." (*Guidelines* @ 5)

- Criteria must be based upon studies showing a clear dose/response relationship to determine effect concentration. Data from confounded studies (i.e., results that are influenced by factors other than the pollutant of concern) should not be used. (*Guidelines* @ 15, 16, 21)

- All decisions should be based on a thorough knowledge of aquatic toxicology and criteria decisions must be altered if there is a substantial probability of over or under protection of aquatic organisms and their uses. (*Guidelines* @ 18)

- Based on "all available laboratory and field information," it must be determined that proposed criteria are "consistent with sound scientific evidence." If not, another criterion should be derived. (The concluding recommendation of the *Guidelines* @ 57)

The basic scientific premise underlying all published EPA nutrient criteria development documents is that nutrient control is intended to reduce excessive plant growth. Consistent with the *Guidelines'* requirements for a clear demonstration of causation, the various EPA nutrient criteria documents for lake and stream environments all clearly specify that dose/response demonstrations are required to set scientifically defensible nutrient standards. Nutrient levels must be documented to "cause" specific changes in plant growth (typically measured as chlorophyll 'a') and other physical variables directly affected by excessive plant growth (Secchi depth, DO, transparency, etc.). See, *Nutrient Criteria Technical Guidance Manual – Rivers and Streams*, USEPA July 2000 (hereafter "*Rivers and Streams Document*"). The *Rivers and Streams Document* is clear that a nutrient criterion must be based on a demonstration that nutrients are causing excessive plant growth (eutrophication). For example, Chapter 1 identifies various ecosystem impacts related to excessive plant growth, and specifies that nutrient criteria are based on the relationship between nutrient levels and plant growth as measured by chlorophyll 'a' ("Nutrient criteria development should relate nutrient concentrations in streams, algal biomass and changes in ecological condition (e.g., nuisance algae accrual rate and deoxygenation). … Initial criteria should be verified and calibrated by comparing criteria in the system of study to nutrients, chl *a* and turbidity values in water bodies of known condition to ensure that the system of interest operates as expected.") *Rivers and Streams Document* @ 13) "Predictive relationships between nutrients and periphyton (or

phytoplankton) biomass are *required* to identify the critical or threshold concentrations that produce nuisance algal biomass." Id @ 76, emphasis supplied.

The various EPA nutrient criteria documents all acknowledge that nutrients may cause ecosystem impacts to upper level organisms (invertebrates, fishes), but *never* directly:

> "However, fish and macroinvertebrates do not directly respond to nutrients, and therefore may not be as sensitive to changes in nutrient concentrations as algal assemblages. It is recommended that relations between biotic integrity of algal assemblages and nutrients be defined and then related to biotic integrity of macroinvertebrate and fish assemblages in a stepwise, mechanistic fashion." *Rivers and Streams Document* @ 85.

EPA's published guidance indicates that invertebrate populations may be affected only when plant growth rises to a level where extensive/excessive plant growth causes those ecosystem changes. These changes are not documented to occur directly due to nutrients as this parameter is not a toxicant and does not have a direct impact on sensitive organisms. See, *Development and Adoption of Nutrient Criteria into Water Quality Standards*, USEPA 2001 @ 14, response 4. This fact was also well documented by EPA's field studies under the whole effluent toxicity program.

The *Guidelines* is quite clear that a simple "weight-of-evidence" approach is not a sufficient basis for setting a criterion. Furthermore, the *Guidelines* provide that a simple regression approach between two variables (one a field response) would not suffice as a demonstration that the input variable caused the effect measured in the field. Nowhere does this document indicate that a conditional probability approach may be used for derivation of a numeric standard. This is not unexpected since that statistical method cannot provide a demonstration that regulating a pollutant at a given level provides any assurance that use protection will or will not be achieved at that pollutant level. At best the method indicates the likelihood (i.e., the probability) of encountering the condition being evaluated for a given pollutant concentration used in the regression. Finally, both the *Guidelines* and the *Rivers and Streams Document* are replete with statements underscoring the need to understand the toxicology of the substance. To set a numeric standard, one must determine that the pollutant of concern is the direct cause of the adverse effect being measured. Simple regressions and conditional probability analyses provide no such confirmation.

Unless there is clear, well documented evidence on the level of water quality necessary and sufficient to protect aquatic life uses, criteria should not be established. Moreover, the *Guidelines* and the *Rivers and Streams Document* are clear that site-specific information should be considered if it shows that the suggested standard is misplaced. The new EPA approach expressly ignores such information. In each case where the new standards were applied in Pennsylvania, it was acknowledged that habitat degradation (sedimentation/channelization) was the root cause of any documented changes in invertebrate populations. Site-specific regressions were provided to demonstrate that, in fact, there was no relationship between nutrient levels and invertebrate populations in the

various streams where such data were available. (Discussed below in greater detail). Contrary to EPA's own recommendations in the *Guidelines* and the *Rivers and Streams Document*, the Region simply ignored those data and analyses, claiming the new procedures provided sufficient confirmation that nutrients were the cause of stream impairments.

Published Literature Confirmed EPA's Simplified Approach to Stream Nutrient Regulation Was Misplaced

As EPA is well aware, the issue of how to develop scientifically defensible stream nutrient standards, in particular, has been a very controversial subject due to the technical challenges inherent in attempting to develop a uniform approach to such waters. It is now well established in the literature that plants inhabiting streams (periphyton/ macrophytes) do not respond as algal species inhabiting lakes (phytoplankton). For example, early on EPA often relied on research produced by Dodds, who sought to develop some type of simplified relationship between periphyton and nutrient levels. (See, *Rivers and Streams Document* @ 77; *Protocol for Development of Nutrient TMDLs, First Edition* – November 1999, USEPA 1999 @ 4-6). However, in Dodds' more recent publications he has concluded that periphyton growth has the capability of reaching very high levels even where very low nutrient levels are present. [1] Thus, it is now apparent that the simplified approaches do not work for controlling periphyton growth in streams as EPA originally had contemplated.

As noted in EPA's criteria documents, in many situations, nutrient levels do not control plant growth but other physical factors do. For periphyton, in particular, this seems to be the case. Likewise, EPA now recognizes that many macrophytes may obtain nutrients from the soil matrix, rendering control of water column nutrient levels a meaningless exercise. (See, *Rivers and Streams Document* @ 73)

The observations of Dr. Dodds were becoming apparent to many others around the country. On July 17, 2007 the Association of State and Interstate Water Pollution Control Administrators (ASIWPCA) sent a letter to the Office of Water informing EPA that states were unable to demonstrate the necessary relationships between nutrient levels and EPA's recommended instream response parameters (e.g., plant growth, turbidity, DO, etc.). (Exhibit 1) That organization raised critical concerns that continuing on the path chosen by EPA would invariably lead to wasted expenditures which state and local governments can ill afford:

> *These problems can only lead to miscues in impairment identification and mis-direction of scarce management and implementation resources.... Because no two water bodies are the same, site-specific evaluations and most probably, site specific criteria are required that reflect their uniqueness and protect their natural trophic tendencies. This will be a costly endeavor but less financially*

[1] Walter K. Dodds. 2006. "Eutrophication and trophic state in rivers and streams." Limnol. Oceanogr. 51(1, part 2) p 671 – 680. "[A]ttached algae might be able to attain impressive biomass in nutrient-poor water because periphyton can use the small amounts of nutrients that continuously flow by.". @ 677

costly than attempting to meet water quality criteria that are unattainable and less environmentally costly than losing water resources because criteria are too liberal.

Thus, ASIWPCA called on EPA Headquarters to reconsider the efficacy of ecoregion or state wide numeric nutrient standards in light of well documented problems in demonstrating that nutrient control would be effective in regulating excessive plant growth in streams. A focus on site-specific conditions was identified as the only way to ensure proper and effective programs for stream restoration. [2]

EPA Region III Mishaps with Nutrient Regulation

EPA Region III itself encountered these same difficulties identified by ASIWPCA in attempting to develop necessary and protective nutrient objectives for five watersheds across the Commonwealth of Pennsylvania. [3] After several years' effort, EPA withdrew the proposed restrictive phosphorus TMDL and instream standards it had developed in late 2006 for Wissahickon Creek. This TMDL sought to limit periphyton growth via point source controls. The site specific periphyton data for that stream, however, clearly documented two facts: (1) upstream of the wastewater plants where TP levels were quite low, periphyton growth was as robust as downstream where TP levels were quite high, and (2) periphyton growth was closely correlated to tree canopy, confirming that light, not nutrients, was the limiting factor for plant growth. [4] EPA coauthored a peer reviewed journal article (with its technical consultant Tetra Tech) that concluded nutrient reduction would not be effective in controlling plant growth because low TP levels could support robust periphyton growth. [5] ("However, it is worth mentioning that while periphyton activities was one of the major causes of the DO violations, it was finally determined to be infeasible to control the periphyton through reducing nutrient loads from point sources.")

Region III Employs an Unprecedented Approach to Nutrient Criteria Development and TMDL Decision Making

Due to the repeated problems encountered in attempting to relate nutrient levels to periphyton growth in streams, the Region employed a new, technically-unprecedented approach to develop stream nutrient standards. The new approach ignored whether nutrient levels affected plant growth and did not even attempt to demonstrate whether nutrients were actually causing any site-specific aquatic life (invertebrate) impairment in

---

[2] Establishing a clear relationship between nutrient loading and plant growth is not generally problematic in lake environments, where such relationships have been well documented for decades.

[3] EPA detailed the problems it encountered in attempting to develop a periphyton/nutrient relationship in a May 2, 2007 affidavit filed by Thomas Henry, USEPA, in the matter of American Littoral Society v. EPA.

[4] Alan Everett. February 19, 2002. Periphyton Standing Crop and Diatom Assemblages in the Wissahickon Watershed. Montgomery and Philadelphia Counties. @ 12

[5] Zou et al,."Integrated Hydrodynamic and Water Quality Modeling System to Support Nutrient Total Maximum Daily Load for Wissahickon Creek, Pennsylvania," Journal of Environmental Engineering, April 2006.

the streams at issue. Rather, the Region acquired phosphorus and invertebrate population data for three "ecoregions" and simply had the data plotted, with total phosphorus or total nitrogen as the independent variable and various invertebrate metrics as the dependent variable. All measured changes in invertebrate populations were assumed to be a direct result of the nutrient concentration exposure.[6] EPA Region III, using the same contractor (Tetra Tech – see footnote 5), developed three "regional" (Eastern Piedmont, Central, and Allegheny Plateau) total phosphorus standards as part of five promulgated nutrient TMDLs in Pennsylvania. The numeric nutrient standards ranged from 25 to 40 ug/l as "growing season" (April to October) averages. (See, *Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania,* Tetra Tech, November 20, 2007; hereafter "*Tetra Tech Report*" and *Development of Nutrient Endpoints for Allegheny Plateau and Ridge and Valley Ecoregions of Pennsylvania: TMDL Application*, Tetra Tech, June 24, 2008.)

In setting these standards, Region III interpreted Pennsylvania's narrative nutrient standard using an approach that deviated from the prior nutrient impact assessment guidance established by Pennsylvania and EPA that required a demonstrated linkage between increased nutrient levels and excessive plant growth.[7] In place of a stream model such as QUAL2 or WASP5, a statistical procedure known as "conditional probability" was used in a "weight-of-evidence" analysis to generate the instream numeric water quality standard. The conditional probability procedure simply plotted datasets from a selected ecoregion that measured instream TP levels and assessed the type of invertebrate populations present. Several different invertebrate indices were chosen as the endpoint of concern (e.g., total taxa, # EPT taxa, % clingers, etc.) for comparison with total phosphorus and total nitrogen levels. A sample of that analysis is presented in Exhibit 2. It is worthy to note that the "weight-of-evidence" analysis evaluated multiple macroinvertebrate metrics that were deemed an important measure of aquatic health and discarded those that did not show sensitivity to TP rather than factor those results into the analysis (See, *Tetra Tech Report* @18).

---

[6] This technical assumption was at odds with prior EPA scientific conclusions regarding the manner in which nutrients affect the environment. EPA's nutrient criteria guidance specifically states that macroinvertebrates ***do not*** respond directly to nutrients. *Nutrient Criteria Technical Guidance Manual – Rivers and Streams,* EPA-822-B-00-0002, Ch. 6, pg. 85 (July 2000).

[7] Pennsylvania DEP's guidance on assessing nutrient impairment to free-flowing streams requires documentation of excessive plant growth caused by increased levels of phosphorus or nitrogen and attendant violations of the dissolved oxygen standard. *Implementation Guidance for Section 95.9 Phosphorus Discharges to Free Flowing Streams*, Department of Environmental Protection Bureau of Watershed Conservation, (October 27, 1997) at pg. 7. ("For purposes of this guidance, a nutrient-related problem is defined as a documented use impairment due to nuisance algal or rooted aquatic plant growth conditions with attendant violations of dissolved oxygen standards."). The designation of a stream as nutrient impaired requires a similar demonstration. *Commonwealth of Pennsylvania Assessment and Listing Methodology*, 2004/2006: *Cause Definitions – Nutrients* ("The presence of excessive quantities of Phosphorus and/or Nitrogen that under the proper conditions may result in dense algal or macrophyte growth and wide fluctuations in Dissolved Oxygen levels.") Regarding the TMDL actions at issue, EPA Region III filed a sworn affidavit with the Court stating that Pennsylvania law required EPA to demonstrate how nutrients impacted plant growth in order to establish a lawful basis to regulate nutrients. (See, n. 3)

Contrary to the *Guidelines* and the *Rivers and Streams Document*, no attempt was made to show that TP was the actual parameter causing the change in invertebrate populations or indices, or to show confounding factors did not influence the invertebrate metrics in EPA's database. Tetra Tech's analysis of the impaired water bodies acknowledged that factors other than phosphorus were causing the changes in invertebrate populations being measured, but EPA had informed Tetra Tech that the waters were nutrient impaired so Tetra Tech's analysis ignored a causal assessment.[8] Tetra Tech also acknowledged that the nutrient levels chosen to protect invertebrate populations *would not* limit plant growth since the selected target was well above published limiting nutrient levels.[9] The method employed was a mere correlation that could not and did not show that nutrients were the cause of the any changes in invertebrate indices.[10]

No attempt was made to demonstrate that the chosen indices were set at a level necessary to protect the stream uses.[11] The acceptable invertebrate levels were simply selected as the midpoint of the rating scale used by the state of Maryland. Based on the selected metric, 50% of the sites used by EPA to evaluate the nutrient standard would be considered impaired or non-attainment sites. Many of these impaired stations have TP levels well below the target TP value established by EPA, confirming that factors other than phosphorus are significantly influencing this database. The most this methodology could indicate is that there was a very weak relationship ($R^2 < 0.1$)[12] between total phosphorus and sensitive invertebrate populations. The R^2 for the data was quite low (~ 0.1 – that is less than 10% of the data response is explained by TP levels). To discern the "protective" standard from this analysis, the contractor looked at the "change point" in the graph. This is the location in the graph where it becomes more probable that the selected instream metric will not be attained. Even at the selected "instream standard" there was a 50% probability that the target invertebrate levels would not be achieved.[13]

---

[8] Response Document for Nutrient and Sediment TMDLs in Pennsylvania for Southampton Creek, Indian Creek, Chester Creek, Paxton Creek, and Sawmill Run, June 30, 2008 @ 14. "Again, this effort was not undertaken to "show" that TP is the cause of impairment … Tt (sic "Tetra Tech") was not asked to determine the cause of impairment; we were given a cause and asked to determine a protective value."

[9] *Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania: TMDL Application*. November 20, 2007. at 15-16. "Not surprisingly, a strong algal biomass-nutrient relationship was not present in our examination of the data sets … Surprisingly, the highest algal biomass occurred as sites where the TP concentrations were relatively low, 14 – 35 ug/L. It is possible that algal growth has been saturated even at this low level."

[10] *Id.* at 11. "Correlation analyses identified significant relationships between biological response and nutrient variables. However, correlation may or may not indicate the real relationship. Numerous relationships were examined; only a subset of which was correlated. There were also results that were considered potentially important but showed weaker relationships (Appendix A)."

[11] In an August 4, 2008 Pennsylvania Right-to-Know Law response, PA DEP stated that it had no documentation or other information showing that the invertebrate target endpoints were necessary or appropriate levels for assessing stream impairments.

[12] *Id* at 18.

[13] Hall and Associates duplicated the Tetra Tech calculations and evaluated the change points associated with varying EPT Taxa conditional probabilities (i.e., <1 – <12 EPT Taxa). The calculations resulted in virtually identical TP concentrations regardless of the conditional EPT Taxa level. To most, this shows that the instream target is not a function of the phosphorus concentration. Incredibly, EPA asserted that this

Nonetheless, both Tetra Tech and EPA concluded that the selected standard would ensure that uses were fully protected and invertebrate impairments restored by meeting this value.

Because EPA's use of the methodology was unprecedented and a radical shift in the established analytical framework, numerous questions were raised in various public meetings held by EPA to present the TMDLs. In these public meetings and subsequent FOIA responses, EPA and its contractor (Tetra Tech) acknowledged that:

- The approach *did not prove* phosphorus was causing the invertebrate response.

- Meeting the chosen numeric nutrient standard would *not ensure* restoration of the target invertebrate population.

- There was no demonstration that the chosen instream metrics were *necessary* to provide use protection; the selected metrics were just the median values from the scoring criteria.

- The approach was not based on any *site-specific information* demonstrating a relationship between elevated TP levels and invertebrate populations.

- The approach did not consider the available site-specific data which, in general, confirmed that factors other than phosphorus were the root cause of the changing invertebrate levels (e.g., habitat alteration).

Regarding the issue of site-specific stream impairment data, EPA applied the new numeric standards to one watershed that was never identified as nutrient impaired on any TMDL list (Chester Creek). The periphyton data provided by EPA as part of the final TMDL confirmed that plant growth in Chester Creek was rather minimal and well below the level EPA thought could cause adverse impacts (i.e., > 150-200 mg/m^2 as a growing season average). Various biological assessments had determined that habitat impairment caused reduced invertebrate populations in one segment of the stream (Goose Creek). This assessment was not considered in applying EPA's new nutrient criteria as the solution to the problem. EPA also ignored the fairly extensive and only site-specific invertebrate data for that watershed, which it had included in the TMDL document. Those data confirmed that phosphorus levels in allegedly impaired segments of the watershed were *unrelated* to invertebrate populations. (Exhibit 3) [14] In fact the *highest* phosphorus levels in the "impaired" segment were associated with the *best* invertebrate

---

complete lack of dose/response confirmed that nutrients are a dramatic stressor. (Supra Note 8 at 17-18) The EPA response is contrary to accepted scientific principles.

[14] EPA's response to comments on the TMDL ignored this analysis finding that the Tetra Tech report was a sufficient basis to conclude that phosphorus was in fact impairing Chester Creek. This position violates the *Guidelines* which requires reconsideration of an approach if the field data confirms it is misplaced (see, *Guidelines* @ 57).

population readings that surpassed the impairment threshold used in the *Tetra Tech Report*. Moreover, the chosen TP standard was violated uniformly in both upstream and downstream waters that DEP had determined fully attained uses. Thus, the new recommended standard applied to Chester Creek does not differentiate between waters with acceptable invertebrate levels and waters with allegedly unacceptable levels.

EPA also applied the new invertebrate impacts-based standard to Paxton Creek. The lower section of this stream that was the focus of the nutrient TMDL is *concrete lined.* Consequently, habitat was well documented as very poor and invertebrate levels were, as expected, quite reduced. Invertebrate levels were robust in other unimpaired segments of the creek, despite having TP levels far in excess of the instream standard that Region III claimed was necessary to protect uses. Analyses of these site-specific data show that the invertebrate populations are much more correlated to habitat and very poorly correlated to phosphorus levels (Exhibits 4 and 5, respectively). As with Chester Creek, the chosen standard (25 ug/l TP growing season average) could not distinguish between impaired and unimpaired segments of the creek. (Exhibit 6) The good correlation between invertebrate population and habitat score confirmed the overwhelming importance of this stressor. However, EPA ignored the ramifications of these data, finding that the new instream TP standard was a valid indicator of impairments due to nutrients.

EPA Headquarters Review Supports The Region's New Approach and Confirms that Nationwide Implementation Of a New Nutrient Criteria Development Procedure is Being Promoted

Given the apparent inconsistency with the *Guidelines* and prior EPA Nutrient Criteria development documents, the inconsistency with the site-specific information, and the complete lack of documentation showing that TP was actually the pollutant causing the changes in invertebrate levels, the group of affected Pennsylvania communities approached EPA Headquarters in April 2008 to request an independent review. Various letters were sent to Benjamin Grumbles and the Office of General Counsel. Initially, the staff's informal response was that the new approach was not consistent with Section 304(a) criteria development requirements and that conditional probability could not be used as the basis to derive a numeric water quality standard. After two months, however, it became apparent that the Office of Water was intent on supporting the new Regional approach to nutrient criteria development using conditional probability and assessing nutrient impacts as if nutrients were toxicants.

To understand the rationale behind this decision, a Freedom of Information Act (FOIA) request was sent to EPA Headquarters to obtain any available background documentation supporting EPA's position. EPA's June 19, 2008 FOIA response confirmed the following. Apparently, after receiving the letter from ASIWPCA, rather than address the difficult technical issues raised (i.e., plant growth in streams is not well connected to nutrient levels), the Office of Water decided to use a new nutrient standard approach that ignored whether or how plant growth was affected by nutrients. A conditional probability approach presented in a paper entitled "*Development of empirical, geographically specific water quality criteria: a conditional probability analysis*

approach" Paul and McDonald (2005), Journal of American Water Resources Association 41:1211-1223 was now the recommended basis for deriving nutrient standards based on invertebrate impacts. (Exhibit 7) [15] In August, 2007 EPA began a nationwide series of "workshops" under the Agency's "N Steps program" to launch its new criteria derivation approach and convince states that this radical new approach was scientifically defensible. Thus, via this series of internal presentations, EPA completely abandoned the published national nutrient criteria approaches specified in the Agency's guidance (e.g., the *Rivers and Streams Document.*) [16] No public notice or peer review of this new approach was given prior to this radical change in nutrient criteria derivation procedures.

It should be noted that Dr. Paul's (EPA ORD) presentation materials entitled "Conditional Probability Analysis: A Statistical Analysis Tool," as well as the original paper co-authored by Dr. Paul (Exhibit 7), contained the following cautions regarding use of conditional probability based upon field data to identify an appropriate instream standard:

> **Disadvantages**: Other stressors confound the association
>
> **Other Points to Remember**
>
> Conditional probability is just a statistical tool that can be used to extract very specific information from a data set. Before applying CPA ("change point analysis"), *it is imperative* that *extensive* laboratory data analysis (EDA) be conducted. EDA is a form of detective work, primarily using graphical depictions of various renditions of the data. (emphasis supplied)

Dr. Paul's examples only suggested applying this procedure in the area of the data where there was great certainty that the stressor was highly correlated to the impairment. EPA left out these cautions in its subsequent September, 2007 Regional Nutrient Criteria Development Workshops.

A June 11, 2008 memorandum by William Sweitlik, Chief, USEPA Ecological and Health Processes Branch, Office of Science and Technology, confirmed that the Regional Office had simply followed EPA Headquarters new advice and should be commended for implementing the procedures in the TMDLs. (Exhibit 8). [17] The memorandum stated the following:

---

[15] In October 2003 the EPA Science Advisory Board considered but did not agree that the conditional probability approach was an appropriate methodology for setting suspended and embedded sediment standards. The procedure was noted as a useful tool, though not sufficient to derive a numeric standard. It was noted that any use of this methodology had to be based on a documented strong stressor response.

[16] The only reference to prior EPA approaches that required a clear demonstration of how nutrients impacted plant growth and then ecosystem indicators was a slide entitled: *Nutrients...ughh.*

[17] William Swietlik to Robert Koroncai. June 11, 2008. Development of Nutrient Endpoints for TMDLs in Pennsylvania.

> It is our conclusion that the approach used in the document (sic –*Tetra Tech Report*) … is a scientifically defensible approach and is consistent with EPA guidance for deriving nutrient criteria. The approach used in the document is an example of the multiple-lines-of-evidence (or weight-of-evidence) approach. … In October 2007 EPA HQ provided training to the Region II and III States on the weight-of-evidence methodology and how it can be applied to developing numeric nutrient values. It is good to see the Region benefited from our training and you are now employing this approach.

Thus, it is apparent that EPA Headquarters has launched a new method for nutrient criteria derivation and that it expects it to be used on a nationwide basis. Discussions with other communities across the country verified that, in fact, state agencies are being requested to use this new approach for standards development.

Among the documents alleged to support the new methods was a prior Science Advisory Board review of the conditional probability method. The development documents specify that the metric for which a criteria is developed using conditional probability must be a strong stressor (i.e., the aquatic community condition is clearly related to the stressor for higher values of the stressor). The background documents cautioned that there must be a clear, scientifically established causal relationship between the pollutant at issue and the endpoint selected for review. [18] The Tetra Tech conditional probability analysis expressly excluded any demonstration that nutrients were, in fact, the cause of any documented change in invertebrate populations. Tetra Tech informed the public that it was directed by EPA to *assume* that the changes in invertebrate populations were caused by phosphorus concentration. [19]

Contrary to the admonitions of Dr. Paul (USEPA ORD) and the Science Advisory Board, the site-specific information for the streams at issue were either unavailable or not considered to determine if TP was a strong stressor or if other factors were confounding the finding that nutrients were the culprit. The site-specific data for Chester Creek and Paxton Creek clearly confirmed TP level was not a strong stressor, if at all. [20] Thus, it is apparent that the Office of Water's

---

[18] The prior consideration of the conditional probability method was for stream sediment impacts. It is well documented that impactedness and sedimentation may severely degrade invertebrate habitats in streams. There is no such scientific demonstration regarding nutrient concentrations. To the contrary, it is well documented that nutrient do not directly impact invertebrates or their habitat.

[19] See, USEPA Region III response to comments on Nutrient TMDLs for Chester, Paxton and Indian Creek. ("Again, this effort was not undertaken to "show" that TP was the cause of impairment. … Tt (sic "Tetra Tech") was not asked to determine the cause of impairment; we were given a cause and asked to determine a protective value.")

[20] The Swietlik Memorandum supported using multiple-lines-of-evidence and cited EPA's support of this approach in guidance documents, including "*Nutrient Criteria Technical Guidance Manual: Rivers and Streams*." In citing this guidance, Headquarters stated that the weight-of-evidence approach combines several approaches including: 1) reference reaches, 2) predictive relationships, and 3) published threshold values. However, the predictive relationships identified by Region III were confounded as standard

support for the Region's approach did not even follow its own guidance on whether such a method may be considered for identifying an appropriate instream standard, when applied to nutrients.

Moreover, while the *Rivers and Streams Document* does discuss considering "multiple-lines-of-evidence" as a way to *strengthen* a scientifically defensible finding, nowhere does that document or the *Guidelines* suggest that a criterion is scientifically defensible simply because it uses "multiple-lines-of-evidence or weight-of-evidence." Finally, nowhere has any published, peer reviewed nutrient criteria development approach stated that it is acceptable to (1) assume impacts are caused by a pollutant, (2) ignore whether plant growth will be affected by nutrient regulation, (3) assume nutrients directly impact invertebrates without documented laboratory studies confirming that fact, or (4) ignore site-specific information that shows nutrient regulation is not necessary or will be ineffective. These new EPA assumptions are radical departures from published, scientifically defensible procedures EPA has used for decades under the Section 304(a) criteria development and 303(d) TMDL programs.

Request for Independent Peer Review

The new approach to developing numeric nutrient standards for streams is scientifically unprecedented and a radical departure from published EPA criteria development methods. If this standards derivation methodology remains unchanged, dischargers throughout Pennsylvania (and eventually the country) will be required to install extremely advanced phosphorus treatment at exorbitant costs with little likelihood of producing demonstrable environmental benefits. While our coalition understands that environmental expenditures will be necessary to ensure that our lakes and rivers meet their designated uses, they are very wary of using their limited resources in an unnecessary fashion or a manner that will not produce the desired results.

EPA Headquarters Office of Water has apparently promoted and now approved the radical new nutrient criteria derivation approach. This new approach has never undergone the peer review or technical evaluation process required of all EPA criteria development changes. For the reasons detailed below, we request that EPA promptly conduct an independent peer review of the new EPA nutrient standard setting approach using either EPA's Science Advisory Board or the National Academy of Sciences.

First, it is apparent that the new approach is contrary to a series of "bedrock" scientific principles relied upon by the Office of Water for decades, including:

---

development tools because "other stressors" exert a greater impact (see *Tetra Tech Report* at 15- 21). As for macroinvertebrates, Region III reported that three of the six metrics considered where either not sensitive to nutrient enrichment or more sensitive to other stressors. Of the remaining three, the regression coefficients were extremely poor. Finally, reference reaches upstream of the municipal facilities confirmed that low invertebrate populations were caused by other stressors, <u>not</u> nutrients. Thus, the primary assumption required for using conditional probability (i.e., the aquatic community condition is clearly related to the stressor for higher values of the stressor) was not met.

- Numeric criteria must be based on documented dose/response relationships between the pollutant and a use impairment (versus assuming the pollutant is causing the problem and ignoring data to the contrary)
- Numeric standards must be set at the level found both necessary and sufficient to protect uses (versus setting the standard where the probability of impacts is decreased even if the stressor response is extremely weak)
- Nutrients are not directly toxic to invertebrates but affect plant growth (versus ignoring the degree of plant growth occurring and assuming that nutrients directly impact invertebrate populations)
- Confounded data may not be used to develop a numeric standard (versus assuming all measured field responses are due to a pollutant, even where the data show this is not true), and
- Site-specific data, when available, must be considered in determining whether a numeric standard is necessary and will achieve its intended level of protection (versus ignoring the site-specific data and assuming that the generalized conditional probability analysis is accurate in all cases)

EPA may not legitimately abandon well established scientific principles and requirements and alter its published criteria development approaches by simply hosting an "ad hoc" series of workshops that recommend that accepted approaches be changed. The public has an absolute right under the Clean Water Act to participate in such critical decision making of nationwide importance. EPA's current approach is contrary to basic principles of administrative law and lacks the transparency that is required of all major regulatory decisions. See, CWA Section 101(e) ("Public participation in the development, revision and enforcement of any regulation, standard, effluent limitation, plan or program established by the Administrator or any State under this Act *shall be provided for*, encouraged and assisted by the Administrator and the States.") There is hardly a more important program than that used to establish the basic water quality criteria for protection of the Nation's waters. Under Section 304(a)(3), ..."Such criteria and information *and revisions thereof, shall be issued to the states and shall be published in the Federal Register and otherwise made available to the public*." The Clean Water Act plainly does not contemplate that major changes to criteria development procedures are to be clandestinely launched via internal EPA workshops and announced to the public as a *fait accompli* under the TMDL program.

Second, federal peer review procedures require that new, innovative or controversial scientific procedures used to establish regulatory program requirements must first undergo peer review before they are used in a regulatory context. EPA has long had a peer review process applicable to changes in criteria derivation methods. (See USEPA Peer Review Policy, 1993) This is a typical situation that would have to undergo federal peer review under EPA's own guidance. In fact, the criteria derivation method employed in this case, and to be employed nationwide, failed to receive Science Advisory Board approval when last considered by that peer review panel. Moreover, on December 16,

2004, the Office of Management and Budget (OMB) issued a final bulletin to all agencies establishing that influential scientific information shall be peer reviewed before it is disseminated by the Federal government. (70 Fed. Reg. 2664, January 14, 2005) EPA updated its own peer review policy to accommodate the OMB requirements (EPA/100/B-06/002, May 2006). Although agencies have discretion to choose the specific type of peer review to employ, the duty to conduct a peer review is not discretionary. *Id.* at 2675. In determining the extent of the peer review necessary, the OMB bulletin stated that "[m]ore rigorous peer review is necessary for information that is based on ***novel methods*** or presents complex challenges for interpretation. Furthermore, the need for rigorous peer review is greater when the information contains ***precedent-setting*** methods or models, presents conclusions that are likely to ***change prevailing practices***, or is likely to affect ***policy decisions*** that have a significant impact." *Id.* at 2668. (emphasis added). There is no serious question that EPA's attempt to use a new scientific approach to nutrient criteria derivation, at odds with its published scientific approach, meets every component of the OMB Bulletin justifying peer review.

Third and finally, the opinions of two internationally renowned experts in the field of nutrient control (who have voluntarily reviewed Region III's approach at our request) state that the approach used and endpoints derived will not ensure designated uses are met, and that using conditional probability to establish the endpoints was not scientifically defensible. (See, letters of Drs. Di Toro and Chapra, Exhibits 9 and 10) [21] These scientists clearly state that EPA's approach is not based on accepted scientific principles and should be peer reviewed by the Science Advisory Board or the National Academy of Sciences, as has occurred in similar cases where new, scientific approaches are being employed in the regulatory program.

As outlined above, the approach used by Region III to set nutrient endpoints in the recently released Pennsylvania TMDLs warrants an independent peer review. Specifically, the new approach is precedent-setting, uses novel methods, and will change prevailing administrative practices. Beyond that, international experts on the issue believe that the approach taken is misguided. Finally, if not modified, the potential cost impact to the Pennsylvania dischargers will be on the scale of billions of dollars, and the nationwide potential to misdirect resources is virtually certain to occur. Misdirection of resources will result in unabated environmental impairments and excess energy usage unrelated to environmental need. The new approach would, in all likelihood, cause more harm than good.

As such we respectfully request that EPA initiate a Science Advisory Board or National Academy of Sciences independent peer review on this new procedure for deriving

---

[21] Dr. Dominic Di Toro and Dr. Stephen Chapra, both internationally recognized experts on environmental pollution matters, state that EPA's approach was (1) not scientifically defensible, (2) did not demonstrate TP was causing any impairment (3) could easily regulate the wrong pollutant (TP instead of sedimentation) and, in any event, (4) did not ensure that the chosen standards would protect the stream uses. Both support that this new procedure should undergo an independent peer review *before* it is used in a regulatory context. The Region had Dr. Di Toro's letter prior to the completion of the TMDL and simply ignored it. However, despite repeated requests, Region III has proffered no credible peer reviewed studies showing that total phosphorus acts like a toxicant and directly impacts sensitive invertebrate populations.

numeric nutrient criteria for streams. Pending that review, it is also respectfully requested that EPA stay further application of this methodology as well as implementation of the TMDLs that were derived using this unauthorized method because EPA failed to follow the statutory requirements of CWA § 304(a) prior to relying on this unorthodox criteria derivation methodology.

EXHIBIT 1

ASIWPCA

Association of State and Interstate
Water Pollution Control Administrators

1221 CONNECTICUT AVENUE N. W, 2ND FLOOR, • WASHINGTON, DC 20036 • TEL: 202-756-0600 • FAX: 202-756-0605 • WWW.ASIWPCA.ORG

Ben Grumbles
Assistant Administrator for Water
US Environmental Protection Agency
1201 Pennsylvania Avenue, N.W.
Room 3219
Washington, DC 20469

July 18, 2007

SUBJECT: Nutrient Pollution and Numeric Water Quality Standards Memo

Dear Mr. Grumbles,

The Association of State and Interstate Water Pollution Control Administrators (ASIWPCA) and its Member States received your memorandum of May 25, 2007, which urges accelerated promulgation of numeric water quality criteria for nutrients and the establishment of additional controls for nutrient pollution. Member States agree that nutrient controls are a critical and necessary component of comprehensive water quality management. In addition to limiting or eliminating discharges of priority pollutants and toxics in treated wastewater effluent, management of water quality is shifting to include controls on less traditional parameters such as sediments and excess nutrients.

However, these parameters also exist as part of a balanced natural aquatic system that is often dynamic and does not exhibit "threshold" effects that are amenable to generic numeric criteria setting. Clearly we are entering a new era of water quality management in which the traditional approach of criteria, permits and enforcement, needs to be reevaluated to ensure its pertinence to Water Quality Standard attainment. This is especially true for nutrients and related response variables that exhibit a wide range of conditions representative of a diversity that must be maintained. The uniformity of eutrophic and productivity conditions that numeric criteria would promote defies both common sense and basic principles of ecological succession, which define homeostasis in the natural world.

The May 25 memo also raises several issues that we believe should be addressed cooperatively by States and EPA. In summary:

- EPA should continue to refine and enhance the scientific basis for numeric nitrogen and phosphorus criteria. Many States are not finding a scientific link between cause and effect that is needed to support numerical standards.
- EPA must not undercut EPA approved State nutrient criteria development plans including their agreed upon milestones and commitments.
- A nutrient criterion requires appropriate implementation procedures. This is a costly endeavor, but well worth it.
- Direct impacts on permittees in the NPDES program need to be considered.

- EPA should support quantitative economic assessment of numerical nutrient controls and management that go beyond local watersheds and State boundaries to include upwind and downstream relationships.
- Given EPA's high level of concern about nutrients, the Agency should develop categorical standards for POTWs and have consistent realistic national effluent limits, rather than relying on a State-by-State battle with poorly supported numerical water quality standards.
- The table in the memo on the status of activities in the States should be updated and clarified.

ASIWPCA strongly encourages EPA continue to refine and enhance the scientific basis for numeric nitrogen and phosphorus criteria, including reevaluating the potential difficulties with statistically-derived generic criteria that may be over or under protective. During their considerable developmental processes, many States are failing to find a strong linkage between the EPA recommended cause variables (N and P) and response variables of chlorophyll-a and transparency, but are finding wide variations in parameters that seem unrelated to professional assessments of "trophic health" status. In many cases, a relationship cannot be demonstrated between causal variables N and P, and factors such as turbidity, light limitation, canopy cover, substrate, aquatic community structure, bioavailability, reservoir sequestration, micronutrient limitations and other "response" variables. These problems can only lead to mis-cues in impairment identification and mis-direction of scarce management and implementation resources.

We emphasize that most States have a "mutually agreed upon" nutrient criteria development plan approved by EPA that includes commitments and specific milestones for the development and promulgation of both narrative and numeric criteria. The May memo seems to override those agreements which States would strongly oppose. While States are actively working to meet the mutually agreed upon milestones of these plans, many are improving water quality through the implementation of their new or existing narrative criteria, often using the TMDL process to meet related criteria, e.g., for dissolved oxygen or aquatic life use support. Nutrient control programs are not waiting on a number and are in fact well underway in every State.

Implementation is a key component of effective water quality management. By itself, the singular act of promulgating a criterion does nothing to improve water quality. A criterion requires appropriate implementation procedures to support its application. Consideration of mixing zones, reasonable potential tests, averaging periods, assessment methodologies, dose-response relationships, and other related factors are all absolutely critical if criteria are to be effectively set and used for developing effluent limits, TMDL targets, non-point source control practices and use impairment decisions for Integrated Reports. Because no two waterbodies are the same, site-specific evaluations and, most probably, site-specific criteria are required that reflect their uniqueness and protect their natural trophic tendencies. This will be a costly endeavor, but less financially costly than attempting to meet water quality criteria that are unattainable, and less environmentally costly than losing water resources because criteria are too liberal.

States are acutely aware that numeric nutrient criteria, whether generic or site-specific, will ultimately have a direct impact on permittees in the regulatory NPDES program, with correspondingly significant monetary impacts. Therefore, in addition to ensuring that the criteria are appropriate to the water body, the States need more capability to quantify the costs and benefits of nutrient controls as they progress toward implementing nutrient control plans. It must be recognized that the practice of promulgating water quality criteria without correspondingly considering implementation is outdated. States also recognize that nutrient pollution in many

Cc: Ephraim King, USEPA
Denise Keehner, USEPA
Grace Robiou, USEPA
Amy Newman, USEPA
Mary Smith, USEPA
ASIWPCA Membership

EXHIBIT 2

Conditional Probability

N-STEPS Objectives

- Provide regions, states, and tribes with support related to nutrient criteria development
- Provide access to expert assistance with issues related to nutrient criteria development and implementation
- Improve communication nationwide

What Is It?

A conditional probability is the probability of an event occurring when some other event has occurred and conditional probability analysis is an approach that allows a user to estimate the likelihood of exceeding some response threshold for a given nutrient concentration. This approach is based on the assumption that as nutrient concentrations increase, the likelihood of an impact on some negative response increases. This approach has great promise in identifying appropriate nutrient criteria to other measures where criteria thresholds have already been developed - for example, dissolved oxygen, pH, or biocriteria.

Example Question: What is the probability of exceeding our dissolved oxygen criterion over a range of different nutrient concentrations and where is the threshold?

How Is It Applied to Nutrient Criteria Development?

Nutrient criteria development involves three main processes: identifying relationships between biological responses and nutrient stressors, examining these relationships, and establishing nutrient and/or biological thresholds or criteria.

Conditional probability analysis (CPA) is statistical tool to evaluate the relative risk of impairment of biological attributes along a nutrient gradient. It calculates the probability of exceeding a given threshold (e.g., biocriterion) given a set nutrient concentration; it can be used, therefore, to identify which nutrient conditions have a significant probability of being associated with adverse biological conditions. By combining CPA with change-point analysis, one can identify the nutrient threshold indicating a high risk of exceeding some other criterion.

How Does It Work?

Conditional probability calculates the probability of an event occurring (e.g., DO<y) when it is known that some other event has occurred (e.g., TP>x). The nutrient concentrations are treated as discrete random variables and probability functions are calculated. Functionally, a two-step procedure is used. One identifies the subset of samples were nutrients exceed some threshold and from those sites, one determines the samples which have exceeded the response threshold. This is the subset of samples for which nutrients exceed some value (x), which are also impacted for the response variable (y). Calculating these values iteratively over the range of nutrient concentrations generated an empirical conditional probability curve. Confidence intervals can be generated for this curve using resampling techniques, like bootstrapping (Paul and McDonald 2005). Thresholds along this curve are identified using variations of change-point analysis (See other fact sheet on change-point analysis).

Data Requirements

Independently collected numeric data in the form of paired observations are required. These are preferably continuous data, although discrete numeric variables (e.g., taxa richness) could also work. The greater the range of environmental conditions encompassed the better. One way to assure a large range is to use a gradient design and select sites along as large a gradient as possible.

What Should You Look For & Report?

Examine the conditional probability plots. They should show a clear trend in response with changes in nutrient concentration. Steep relationships and clear changes in the response make threshold identification easier. Confidence intervals can be used to help identify thresholds (see Paul and McDonald 2005). Confidence intervals for the change-points can also be reported (see fact sheet on Change-Point Analysis), depending on the approach used.

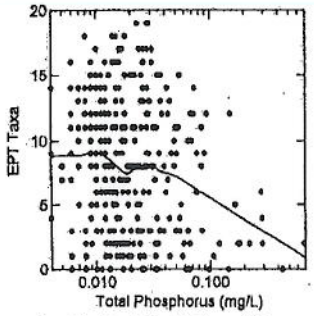
Pros	Cons
• Effective way to identify nutrient thresholds	• Requires substantial data, probabilistic data is best
• Nice approach for linking criteria together	• Lack of significance does not mean lack of association
• Quantitative measure of thresholds with error estimate	• Requires knowledge of an appropriate threshold for response variable
	• Lack of significance does not mean lack of association

Alternatives

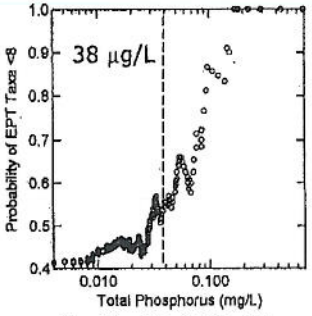
- Change-point analysis alone

Citations

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



Scatterplot with Loess Line



Conditional Probability Plot

For more information contact:
Steve Potts
US EPA
202-566-1121
Potts.Steve@epa.gov

EXHIBIT 3

Chester Creek TMDL
Summary of Macroinvertebrate Metrics and Nutrient Concentrations in East Branch Chester Creek (1996/1997)

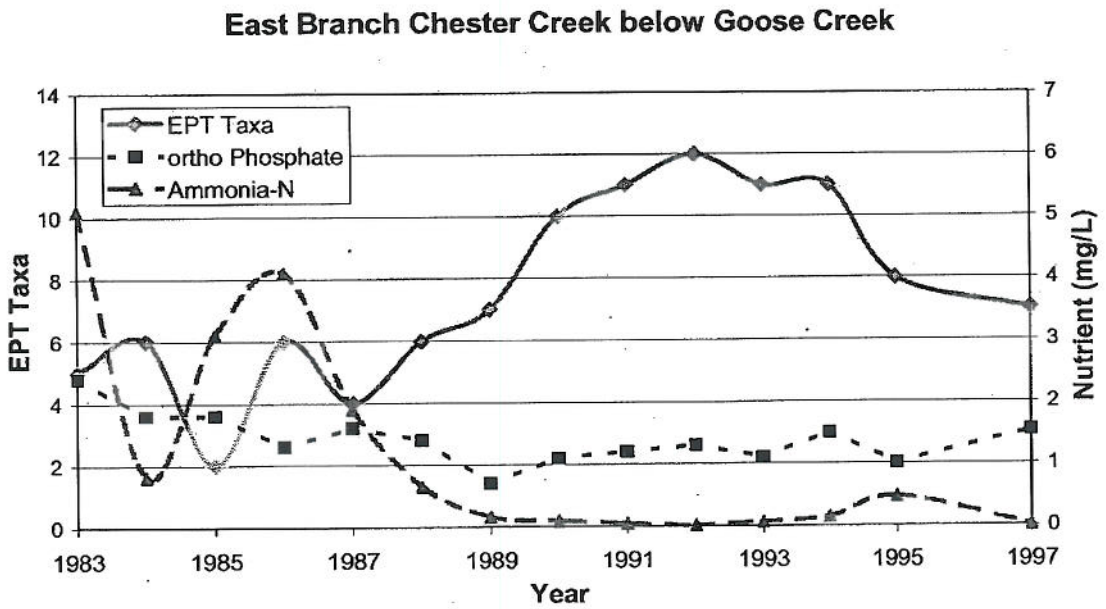
Station	Year	Total Taxa [1]	Total EPT Taxa [1]	HBI [1]	TN [2] (mg/L)	TP [2] (mg/L)
1476790	1997	17	9	4.2	4.3	0.018
1476830	1996	29	10	5.6	2.4	0.020
1476835	1996/1997	30/26	13/12	4.7/4.6	3.0	0.205

Definitions: EPT – Ephemeroptera, Plecoptera, and Trichoptera; HBI – Hilsenhoff Biotic Index; TN – total nitrogen; TP – total phosphorus

[1] Macroinvertebrate metrics from USGS Water-Resources Investigations Report 02-4242 "Assessment of Stream Conditions and Trends in Biological and Water-Chemistry Data from Selected Streams in Chester County, Pennsylvania, 1981-97" by Andrew G. Reif @ 60-62.

[2] Nutrient averages from Table 3-18, Nutrient Total maximum Daily Loads for the Chester Creek Watershed, Pennsylvania – Draft Report, February 2008.

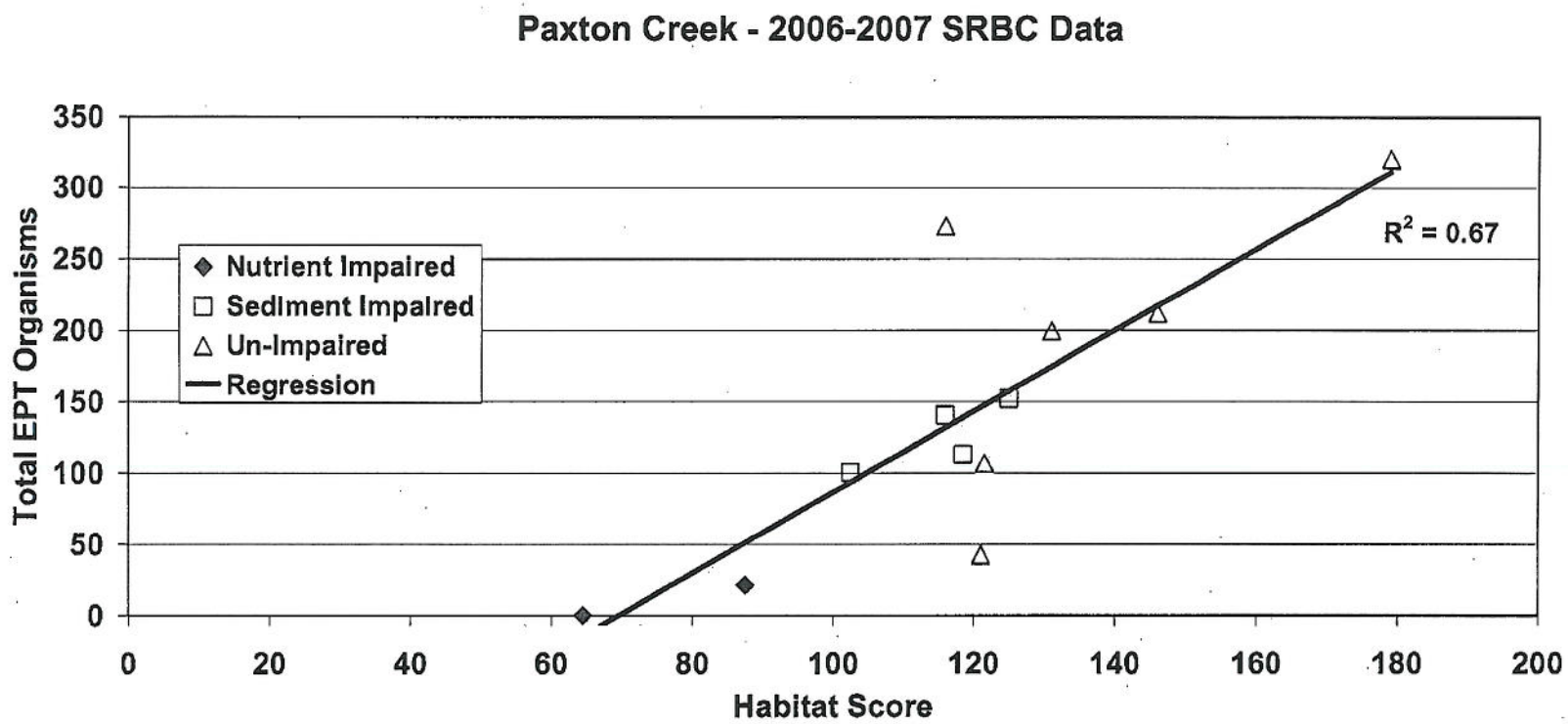
Temporal Changes in EPT Taxa, Ammonia-nitrogen, and Ortho-phosphate at USGS Station No. 01476848 (East Branch Chester Creek below Goose Creek near West Chester)



Macroinvertebrate metrics from USGS Water-Resources Investigations Report 02-4242 "Assessment of Stream Conditions and Trends in Biological and Water-Chemistry Data from Selected Streams in Chester County, Pennsylvania, 1981-97" by Andrew G. Reif @ 62.

Analytical data for ortho-Phosphate and Ammonia-nitrogen from USGS Open-File Report 99-216 "Physical, Chemical, and Biological Data for Selected Streams in Chester County, Pennsylvania, 1981-94 by Andrew G. Reif @ 127-128 and USGS Open-File Report 00-238 "Physical, Chemical, and Biological Data for Selected Streams in Chester County, Pennsylvania, 1995-97 by Andrew G. Reif @ 11.

EXHIBIT 4

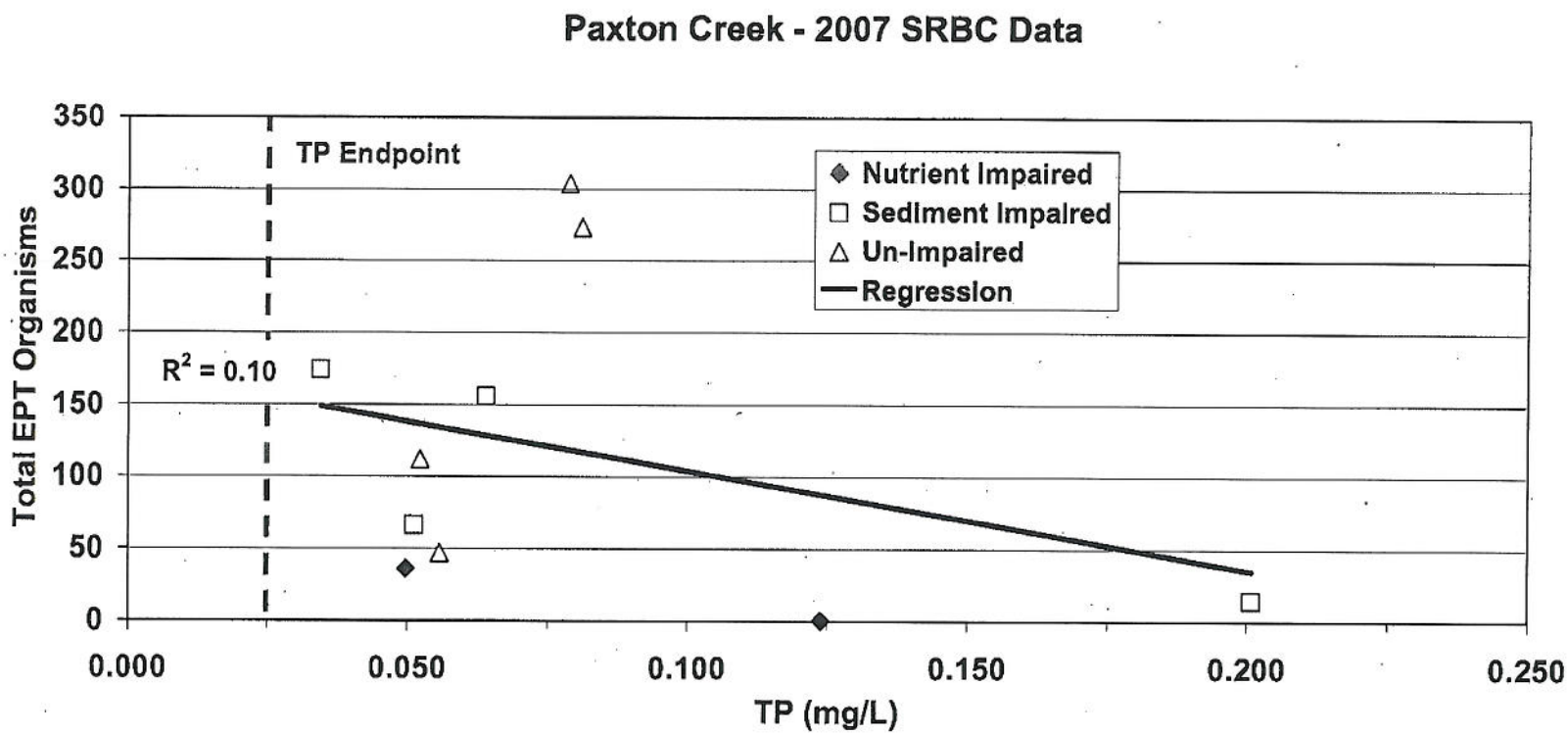


EPT – Ephemeroptera (Mayflies), Plecoptera (Stoneflies), and Trichoptera (Caddisflies)

Habitat scores from Nutrient and Sediment Total Maximum Daily Load in Paxton Creek Watershed, Pennsylvania (June 30, 2008) Table A-4 @ A – 23.

Macroinvertebrate metrics data (total EPT taxa) from Nutrient and Sediment Total Maximum Daily Load in Paxton Creek Watershed, Pennsylvania (June 30, 2008) Table A-5 @ A-25 and Table A-6 @ A-26.

EXHIBIT 5

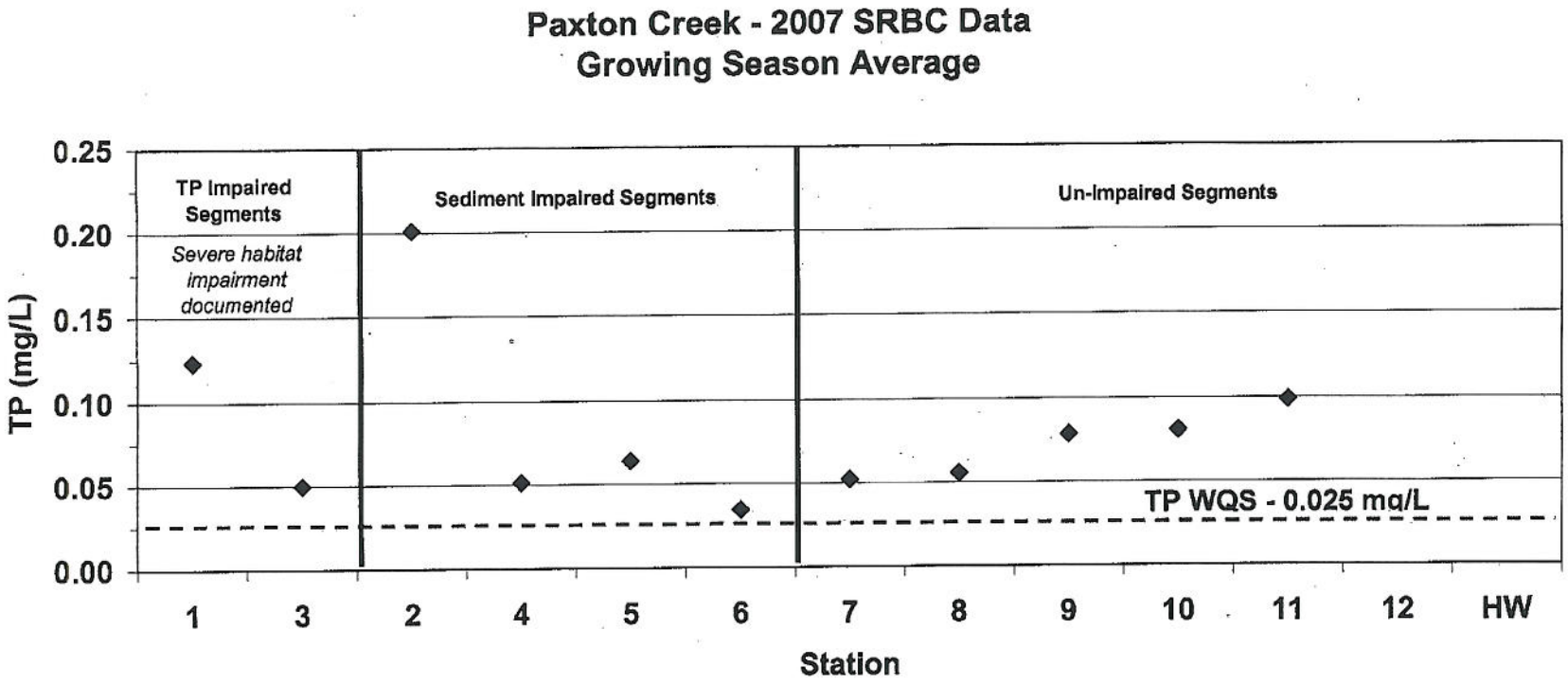


EPT – Ephemeroptera (Mayflies), Plecoptera (Stoneflies), and Trichoptera (Caddisflies)
TP – total phosphorus

Nutrient concentration data obtained from Response Document for Nutrient and Sediment TMDLs in Pennsylvania for Southampton Creek, Indian Creek, Chester Creek, Paxton Creek and Sawmill Run. June 30, 2008. @ 89 - 91

Macroinvertebrate metrics data (total EPT taxa) from Nutrient and Sediment Total Maximum Daily Load in Paxton Creek Watershed, Pennsylvania (June 30, 2008) Table A-6 @ A-26.

EXHIBIT 6



TP – total phosphorus
WQS – water quality standard

Nutrient concentration data obtained from Response Document for Nutrient and Sediment TMDLs in Pennsylvania for Southampton Creek, Indian Creek, Chester Creek, Paxton Creek and Sawmill Run. June 30, 2008. @ 89 - 91

EXHIBIT 7

JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
OCTOBER AMERICAN WATER RESOURCES ASSOCIATION 2005

DEVELOPMENT OF EMPIRICAL, GEOGRAPHICALLY SPECIFIC WATER QUALITY CRITERIA: A CONDITIONAL PROBABILITY ANALYSIS APPROACH[1]

John F. Paul and Michael E. McDonald[2]

ABSTRACT: The need for scientifically defensible water quality standards for nonpoint source pollution control continues to be a pressing environmental issue. The probability of impact at differing levels of nonpoint source pollution was determined using the biological response of instream organisms empirically obtained from a statistical survey. A conditional probability analysis was used to calculate a biological threshold of impact as a function of the likelihood of exceeding a given value of pollution metric for a specified geographic area. Uncertainty and natural variability were inherently incorporated into the analysis through the use of data from a probabilistic survey. Data from wadable streams in the mid-Atlantic area of the U.S. were used to demonstrate the approach. Benthic macroinvertebrate community index values (EPT taxa richness) were used to identify impacted stream communities. Percent fines in substrate (silt/clay fraction, < 0.06 mm) were used as a surrogate indicator for sedimentation. Thresholds of impact due to sedimentation were identified by three different techniques, and were in the range of 12 to 15 percent fines. These values were consistent with existing literature from laboratory and field studies on the impact of sediments on aquatic life in freshwater streams. All results were different from values determined from current regulatory guidance. Finally, it was illustrated how these thresholds could be used to develop criterion for protection of aquatic life in streams. (KEY TERMS: sediment; wadable streams; benthic community condition; statistical analysis; aquatic ecosystems; standards.)

Paul, John F. and Michael E. McDonald, 2005. Development of Empirical, Geographically Specific Water Quality Criteria: A Conditional Probability Analysis Approach. Journal of the American Water Resources Association (JAWRA) 41(5):1211-1223.

INTRODUCTION

A range of procedures are being used around the world for establishing criteria for the protection of water quality (Jimenez *et al.*, 1999; Yin *et al.*, 2003; Borja *et al.*, 2004; Kamizoulis and Saliba 2004, Kay *et al.*, 2004). In the United States, the U.S. Environmental Protection Agency (USEPA) is responsible for implementing the Clean Water Act (CWA) (Russo, 2002), which is the major national act for protecting water quality. The USEPA implements some aspects of the CWA by providing guidance for the control of pollutants through development of Water Quality Standards (WQS). These WQS serve as the foundation for pollution control and are a fundamental component of water quality management. They define the goals for a waterbody by designating its uses, setting criteria to protect those uses, and protecting water quality through antidegradation provisions. The criteria are developed for the protection of aquatic life as well as for human health.

Water quality criteria (WQC) for individual chemical pollutants (such as heavy metals and synthetic organic compounds) have been developed as national criteria (e.g., USEPA, 1994). These national criteria have been developed through laboratory bioassays, where exposure to a single pollutant can be maintained under controlled conditions (Hohreiter and Rigg, 2001; Rausina *et al.*, 2002; Fisher and Burton 2003). As progress has been made in controlling these individual pollutants, a shift has occurred toward control of nonpoint source pollution (e.g., runoff, nutrients, and sedimentation). Consistent with this is the increased use of biological indicators to assess the condition of the environment (Niemi and McDonald, 2004). Recent developments suggest that adopting national criteria may not be sufficiently protective of the biota in various subregions (Perry and Vanderklein, 1996; USEPA, 2000a), thus leading to a greater reliance on field generated data. The recent development of WQC for nutrients is an example of

[1]Paper No. 04095 of the *Journal of the American Water Resources Association* (JAWRA) (Copyright © 2005). **Discussions are open until April 1, 2006.**

[2]Respectively, Research Environmental Scientist and Director, Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Mail Drop 343-06, Research Triangle Park, North Carolina 27711 (E-Mail/Paul: Paul.john@epa.gov).

geographically specific criteria developed from field data (USEPA, 2000a).

Sedimentation in streams is an example of a nonpoint source pollution problem (Spooner *et al.,* 1991). Excessive sediment is a major cause of impairment in waterbodies across the country (USEPA, 2002). Development of water quality criteria for suspended and bedded sediments for the protection of aquatic life provides a challenge since the traditional approach using laboratory bioassays (dose-response studies) may not be applicable (Perry and Vanderklein, 1996).

A focus on the response of aquatic communities to sedimentation emphasizes protection of these communities from adverse effects of sedimentation and is consistent with the use of biological criteria. Establishing a criterion for sedimentation allows the source of excess sediments to be addressed for regulation or remediation of the problem. However, identification of the source of the sediments is not necessary for the development of criteria and is not discussed further.

In this paper, it is shown how a conditional probability analysis can be used with empirical, probabilistic monitoring data for aquatic resources to establish thresholds of impact for a stressor for a specified geographic area. Scientifically defensible thresholds are a necessary first step in establishing protective criteria by environmental managers. The approach is demonstrated by applying it to wadable streams in the mid-Atlantic region of the United States, and establishing thresholds for impact of sedimentation on the streams in this region. Finally, an illustration is presented on how these thresholds could be used to develop a criterion for protection of aquatic life in these streams from sedimentation.

DATA SOURCES

The field data used in this paper are available through the USEPA Environmental Monitoring and Assessment Program (EMAP) web site (USEPA, 2004). These data were collected from the mid-Atlantic region streams in 1993 and 1994 and include

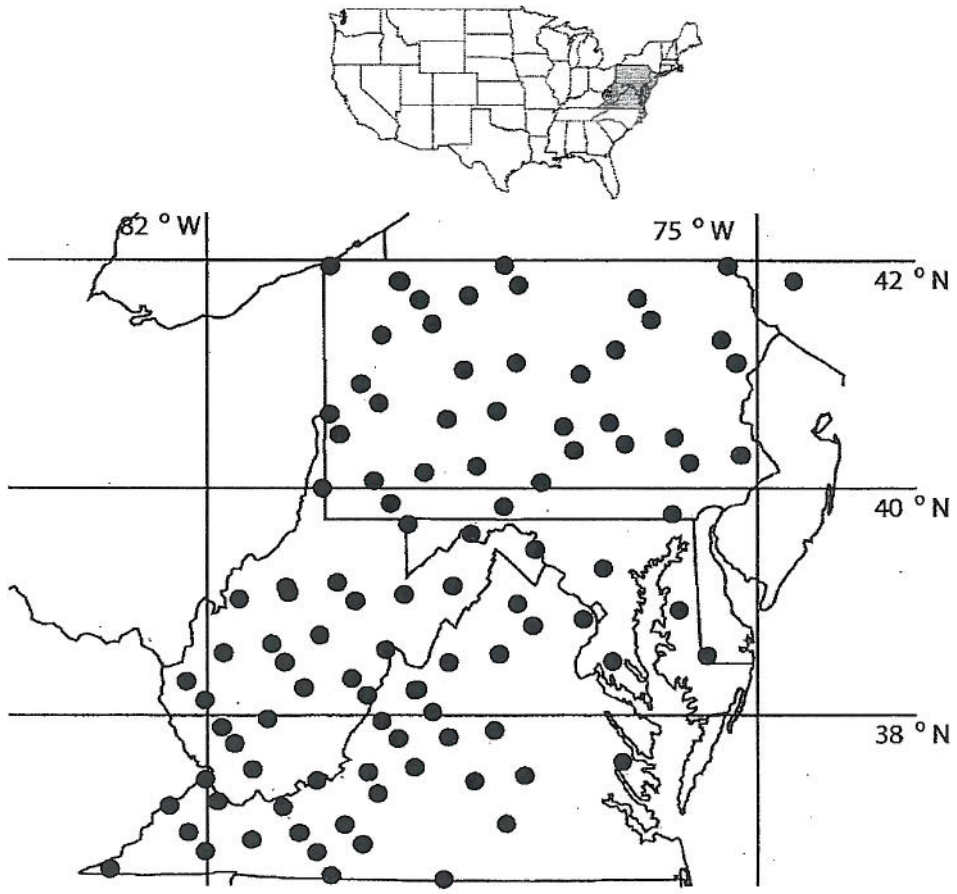


Figure 1. Mid-Atlantic Region of the U.S. With the Wadable Stream Sampling Sites Used in This Study.

102 stream segments in first to third (Strahler) order wadable streams (Figure 1). These segments were selected for sampling using a spatially balanced probability design (Stevens, 1997; Stevens and Olsen, 1999). Inclusion probabilities for each sampled stream segment were determined using the sample sizes for each Strahler order and the total length of streams within each order in the region. Sampling locations within stream segments were chosen randomly. Quantitative data for stream macroinvertebrates, habitat, and water quality were collected at each site (for specifics see Lazorchak *et al.*, 1998; Kaufman and Robinson, 1998; Herlihy *et al.*, 2000; and Klemm *et al.*, 2002). Sampling took place during a two-month sampling window each year from April through mid-June.

Stream Macroinvertebrate Data

Stream benthic macroinvertebrates are a robust measure of stream condition, integrating temporal pollutant exposure. They are responsive to changes in in-stream sediment levels (Davis and Lathrop, 1992; Covich, 1999). Benthic stream community taxa in the orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) (collectively known as EPT) are considered reasonably sensitive indicator organisms since they exhibit a decrease in taxa richness with increased degradation of stream conditions (Loch *et al.*, 1996; Barbour *et al.*, 1999; Zweig and Rabeni, 2001; Klemm *et al.*, 2002; Kaller and Hartman, 2004). The EPT taxa were used to identify impacted stream segments in the mid-Atlantic; when EPT taxa were less than 9 in the stream segments, the stream segments were considered to be impacted (Davis and Scott, 2000; Klemm *et al.*, 2002).

Indicator for Sedimentation

Sediments, including suspended and bedded, can directly affect stream biota or indirectly affect stream biota through changes in habitat. For example, excessive suspended sediments in aquatic systems can cause increased turbidity and decreased light penetration. Altered light regimes can directly alter primary productivity and increase shading of submerged macrophytes (Canfield *et al.*, 1985; Best *et al.*, 2001). Excess fine sediments can fill in gaps between larger substrate particles, embedding the larger particles and eliminating interstitial spaces that would otherwise be used as habitat for reproduction, feeding, and cover for invertebrates and fish (Suttle *et al.*, 2004). For example, bedded sediments in streams and rivers can cause the loss of spawning habitat for salmonids due to increased embeddedness (Young *et al.*, 1991).

For the purpose of this paper, percent fines in the substrate is used as a surrogate indicator for sedimentation in streams. The percent fines (silt/clay fraction, less than 0.06 mm) is a direct measure of the smallest class of sediments and is strongly correlated with sediment embeddedness, a source of the most likely to be resuspended sediment, and an indirect measure of suspended sediment levels in the water column. Streams containing a larger fraction of fine sediment would be expected to have a benthic community at greater risk for impact (Zweig and Rabeni, 2001). Details of the protocols for sample collection and analysis for percent fines can be found in Lazorchak *et al.* (1998), Kaufmann *et al.* (1999), and Klemm *et al.* (2002). Percent fines is determined by visual examination at 11 equally spaced stream transects. The data used were restricted to stream segments with pools.

Reference Conditions

Reference conditions are expectations as to the condition of biological communities in the absence of any human disturbance (Plafkin *et al.*, 1989; Gerritsen *et al.*, 1994). These conditions provide an estimate of natural variability in biological condition and habitat quality that can be expected to occur. Few streams in the mid-Atlantic area are undisturbed. Reference conditions in the mid-Atlantic have been identified (Waite *et al.*, 2000) using a set of selected chemical and habitat conditions from unimpacted or minimally impacted streams in the area. These chemical and habitat measures are used to identify the best available biological conditions in the area streams.

The chemical and habitat parameters used in the selection of reference conditions for wadable streams in the mid-Atlantic area (Waite *et al.*, 2000) are: acid neutralizing capacity (> 50 μeq/L), chloride (< 100 μeq/l), sulfate (< 400 μeq/l), total nitrogen (< 750 μg/l), total phosphorous (< 20 μg/l), and mean rapid bioassessment protocol (RBP) habitat score (> 15). The RBP habitat score encompasses the variety and quality of the substrate, channel morphology, bank structure, and riparian vegetation. It is measured on a scale of 1 to 20, where 1 is very poor habitat and 20 is excellent habitat (Barbour *et al.*, 1999). Waite *et al.* (2000) define a stream segment as reference if all six of the chemical and habitat parameters are within the desired levels. Reference sites were identified that met these criteria for the extant data. Thus, the number and distribution of reference site conditions were not predetermined and were used strictly for comparative purposes.

METHODS

Conditional Probability Analysis (CPA)

The probability of observing a certain event y is denoted P(y). Data acquired with a probability survey design provide estimates of the probability of occurrence for a sampled variable. For example, consider a sampling frame that includes all stream segments in a state. If 75 percent of the sampled stream segments exhibit impacted benthic communities, then the likelihood of observing benthic impact in any of the stream segments in the state is 75 percent.

For use in developing a numeric water quality criterion, a conditional probability statement provides the likelihood (probability) of impact, if the value of a pollution metric (threshold) is exceeded. Conditional probability is the probability of an event when it is known that some other event has occurred, and is denoted

$$P(y \mid x^*) \tag{1}$$

where y is the event of interest and x^* is the other event that has occurred previously. For criterion development, x^* is replaced with $x > x_C$, where x_C is the specific threshold that is exceeded. Therefore, the conditional probability statement in this paper is the probability of an event y occurring, when it is known that some event x has occurred and x has exceeded some threshold x_C [$P(y \mid x > x_C)$]. For water quality criteria, this implicitly assumes that as the pollution metric of interest increases, the likelihood of an impact on biological condition increases. For responses that increase as the pollution metric decreases (e.g., dissolved oxygen concentration), the conditional probability description is reversed [$P(y \mid x < x_C)$].

The pollution metric, x, is treated as a discrete random variable. The probability mass function (pmf) of x (the probability that x is a specific value) is represented by $f_i(x_i)$, where i is 1,2, ... N, and N is the total number of samples. The inclusion probability provides the probability for selection of a given sample, that it can be different for each sample, and is used as a weighting factor for statistical estimation. Every sampling unit has a nonzero probability of being selected. The sum of f_i over all possible sample sites is 1.

It is assumed that an appropriate response variable for stream condition, y_i, exists, which also is a discrete random variable, and equals 0 for good conditions and 1 for impacted conditions. The y_i will be paired with values of the pollution metric, x_i. The value of y_i will be dichotomous (0,1) depending on the value of x_i and can be written as a function of x_i, $y_i = g(x_i)$. The conditional probability, in this work, of impacted conditions if specific value, x_C, is exceeded, is

$$P(\mathbf{y} = 1 \mid \mathbf{x} > \mathbf{x}_C) = \frac{P(\mathbf{y} = 1, \mathbf{x} > \mathbf{x}_C)}{P(\mathbf{x} > \mathbf{x}_C)} \tag{2}$$

where $P(y = 1, x > x_C)$ is probability of joint occurrence for the two events. This equation is the definition of conditional probability (Hogg and Ledolter, 1992). The conditional probability can be expressed as

$$P(\mathbf{y} = 1 \mid \mathbf{x} > \mathbf{x}_C) = \frac{\sum_{x_i > x_c} g(x_i) f_i(x_i)}{\sum_{x_i > x_c} f_i(x_i)} \tag{3}$$

Functionally, to determine the probability of impact in the stream when some value of the pollution metric, x_C, is exceeded, $P(y = 1 \mid x > x_C)$, a two-step procedure is used. Using the survey data, one identifies a subset of the sampled resource (e.g., stream segments) for which $x > x_C$ (i.e., the stream segments are stratified based on the value of the pollution metric x); and from this subset of stream segments in which $x > x_C$, one determines those segments in which the biological conditions are impacted. This is the subset of the sampled stream segments in which the pollution metric exceeds a specific value (x_C), and which are also biologically impacted.

The probability of the biology being impacted in stream segments when $x > x_C$ over the entire range of observed x provides an empirical conditional probability curve. Confidence intervals (CIs) for this empirical curve can be estimated by bootstrap resampling (Manly, 1997). Bootstrapping assumes that the distribution of a population can be determined by resampling the original data. A bootstrap sample consists of drawing a sample of size N from the original data (of size N) with replacement, which is then used to calculate a bootstrap value for conditional probability, $P(y = 1 \mid x > x_C)$. One thousand samples were generated for the bootstrap distribution. The 90 percent and 95 percent CIs were determined from the empirical percentiles (Insightful Corp., 2001).

Identifying Thresholds of Impact

Threshold levels for pollutants that elicit different levels of biological impact in stream segments of a region need to be identified for eventual use in developing criteria. A threshold of impact was identified as

a changepoint separating the empirical conditional probability curve into two parts, that part of the curve above the changepoint and that which is below it. For those samples that are above the changepoint, the probability of impact is different from what one would expect for the entire geographic area. A confounding factor in the identification of a changepoint is that these two groups created by the changepoint are not independent (i.e., the numbers used to create the points above the changepoint are a subset of the numbers used to create the points below the changepoint). Thus, a traditional t-test cannot be used in the determination of the changepoint since the data are not independent (Venables and Ripley, 1997). The identification of the changepoint was by using a weight-of-evidence approach with three different techniques. These techniques are: nonoverlapping confidence intervals, change in curvature of fitted curve, and nonparametric deviance reduction. Other possible techniques could be used to identify a changepoint. In this demonstration, specific values for factors and CIs were selected only as examples. Values used in an actual application of this approach would depend on the particular management requirements and objectives.

The use of nonoverlapping CIs to determine a changepoint involves determining when the lower CI of the empirical curve no longer overlaps the upper CI of the unconditional value (Cherry, 1996; Rahlfs, 1997; Cherry, 1998; Austin and Hux, 2002). This procedure is a conservative estimate for significant difference, since the CIs could overlap when the values are significantly different (Austin and Hux, 2002). The bootstrap percentile CIs based on a bootstrap distribution of 1,000 samples were used for this evaluation. The α-level for the nonoverlapping CI must be adjusted to account for the one-sided nature of this test, whereas the α-level for developing the CIs for the curves was based on a two-sided test (i.e., a factor of 2 in the α-level).

The second technique used for selecting a threshold of impact through changepoint identification was to fit an equation to the empirical curve for conditional probability. The following constraints were used: the conditional probability approaches the unconditional value, $P(y = 1)$, as x goes to the minimum x-value; the conditional probability approaches 1 as x goes to the maximum value; and there is a curvature change at the inflection point of the curve. The following functional form satisfies these constraints

$$P(y=1|x>x_C)=\begin{cases}1+(D_0-1)/\left(1+\exp(B_0(x_C-x_0))\right), \text{ for } x_C>x_0\\ 1+(D_0-1)/\left(1+\exp(B_1(x_C-x_0))\right), \text{ for } x_C\le x_0\end{cases} \quad (4)$$

where exp is the exponential function to base e, D_0 is unconditional probability value $P(y = 1)$, x_0 is the changepoint where curvature changes, B_0 is curvature for values of $x_C > x_0$, and B_1 is curvature for values of $x_C \le x_0$. The parameters x_0, B_0, and B_1 are determined from a nonlinear least squares regression (Venables and Ripley, 1997). Uncertainty in the parameters are estimated from the standard errors generated by the regression software and, where possible, by computing asymmetric confidence intervals (Venables and Ripley, 1997). The residuals from the regression were checked for normality. While it may be generally possible to fit Equation (4) to the empirical curve, the curvature values (B_0 and B_1) may not be significantly different, and a threshold would not be identified with this technique.

The third technique uses nonparametric deviance reduction to determine the changepoint. This approach determines the dividing point for splitting the data into two groups, resulting in the largest reduction in the deviance in the data (Qian *et al.*, 2003). The deviance is defined as

$$D=\sum_{i=1}^{N}(P_i-P^*)^2 \quad (5)$$

where D is the deviance, N is the sample size, Pi is the conditional probability $P(y = 1 \mid x > x_i)$, and P^* is the mean of P_i based on a sample size of N. When the data are divided into two groups, the sum of the deviance for the two subgroups is always less than or equal to the deviance for the entire data set. When the split in the data minimizes the deviance, the threshold is identified. This approach has been used to detect ecological changes along an environmental gradient (Qian *et al.*, 2003). Qian *et al.* (2003) compared results of deviance reduction with a Bayesian hierarchical modeling approach and found that the nonparametric approach provides similar results with the Bayesian analysis.

The deviance reduction point generally can be determined, but it may or may not be of biological significance. Uncertainty in the deviance reduction changepoint (90 percent and 95 percent CIs) is estimated from the empirical percentiles for the bootstrap distribution from resampling 1,000 times (Manly, 1997). An approximate χ^2 test was used to determine the significance of the changepoint. The test assumes that the deviance reduction divided by the scale

parameter is approximately χ^2 distributed with 1 degree of freedom (Venables and Ripley 1997). A large deviance reduction will result in a small p-value, and the consequent rejection of the null hypothesis (H_0: no changepoint).

Biological Importance of Identified Thresholds

For use in criteria development, some level of biological importance needs to be associated with the threshold of impact value that is identified. The changepoint value determined by each technique must separate the samples so that the probability of impact for samples above the threshold would be different from what one would expect for the entire geographic area. A summary of literature values on the response of fish and benthic invertebrates at low reported levels of percent fines in the substrate (Newcombe and Jensen, 1996; Bash *et al.*, 2001; Berry *et al.*, 2003) was used to identify biological importance.

Statistical Analysis of Data

The cumulative distribution function (CDF), the conditional cumulative distribution function (CCDF), and their reverses were used to supplement the conditional probability analysis. The CDF gives probability that x is less than or equal to x_C

$$P(x \le x_C) \approx F(x_C) = \sum_{x_i \le x_c} f_i(x_i) \tag{6}$$

The reverse CDF is the probability that x is greater than x_C, which is the complement of Equation (6), or

$$P(x > x_C) \approx 1 - F(x_C) = 1 - \sum_{x_i \le x_c} f_i(x_i) = \sum_{x_i > x_c} f_i(x_i) \tag{7}$$

The CCDF is the distribution for a subset of the total data, subsetted by (or conditioned on) a second variable [F(y | x)]. The reverse CCDF is similar to Equation (7), that is, 1-F(y | x). The reverse functions are consistent with the CPA results, which are expressed as a threshold (i.e., exceeding some value x_C).

RESULTS

The CDF for EPT taxa richness is shown in Figure 2. Approximately 42 percent of the stream miles across the region were observed to have EPT taxa

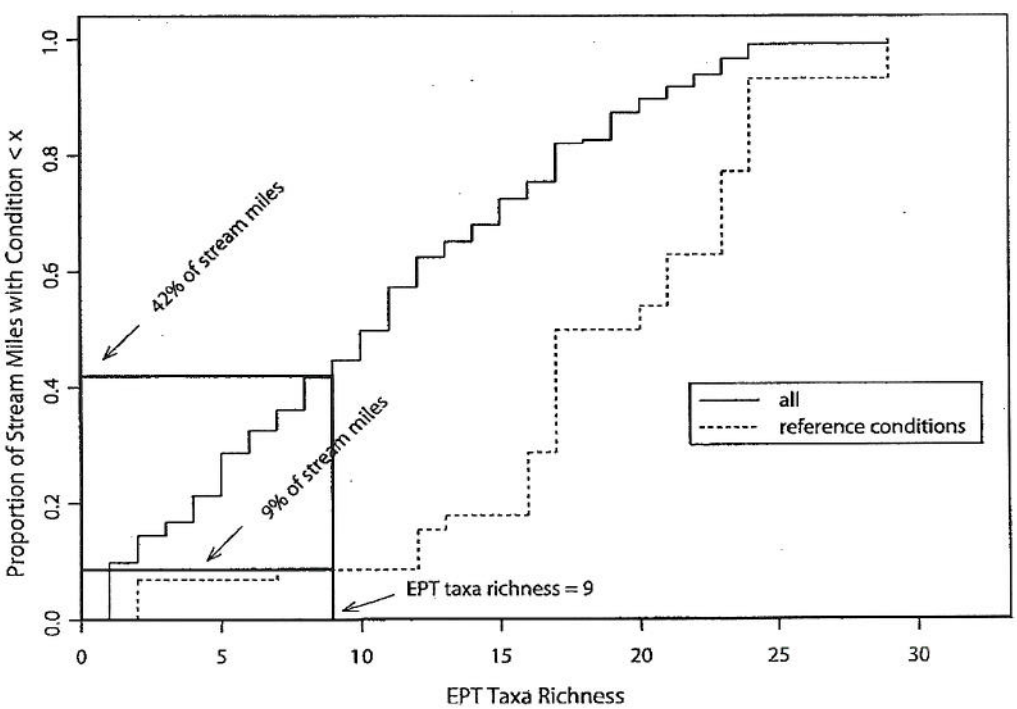


Figure 2. Cumulative Distribution Function for EPT Taxa Richness for All Stream Miles and for Stream Miles That Exhibit Reference Condition Characteristics.

richness less than 9, indicating impacted benthic community conditions. Out of 100 stream segments that had valid values for the indicators used in this study, 16 met the reference condition requirements. In 91 percent of the reference condition stream miles, benthic communities were found that had EPT taxa equal to or greater than 9 (Figure 2). The EPT taxa richness generally declines as the percent fines increases (Figure 3, correlation coefficient, r, is -0.50). The fraction of EPT taxa richness variance explained using a linear regression with percent fines as the predictor is 0.25, suggesting that percent fines does appear to have a substantial effect on EPT taxa richness.

The reverse CDF and reverse CCDFs for percent fines in the substrate were expressed as a proportion of stream miles (Figure 4). The sampled stream segment values were weighted by inclusion probabilities to convert to stream miles. The distribution for impacted benthic communities is displaced to the right of that for benthic communities in good condition (Figure 4). The distribution for reference conditions is shifted to the left (towards lower percent fines) than that for unimpacted streams (Figure 4), since these are the best observed conditions.

The CPA approach suggests that when percent fines in the substrate is greater than 49 percent, there is a 100 percent probability that the benthic communities are impacted (Figure 5). All sites with percent fines in the substrate in excess of 49 percent had EPT taxa richness less than 9. As the percent fines approaches zero, there is a background level of impact on EPT taxa richness from all sources of stress in the region (mean = 42 percent, 95 percent confidence interval of 30 to 56 percent). Thus, irrespective of the level of percent fines in the substrate, approximately 42 percent of the stream miles in the region will likely exhibit an impact on EPT taxa richness. Therefore, to detect a significant signal due to percent fines in the substrate affecting the EPT taxa richness, the upper confidence limit on the estimate of the background impact (e.g., 56 percent, Figure 5) must not overlap with the lower confidence limit on the probability of benthic impact curve in Figure 5. The point at which this occurs is when the percent fines in the substrate is 14.8 percent (Figure 5). This is a threshold of impact, and is statistically distinguishable from background within this geographic area. The mean probability of observing impacted EPT taxa richness associated with this threshold is 67 percent.

The CPA identified threshold of 14.8 percent fines (from nonoverlapping CIs) would translate into approximately 47 percent of the total stream miles in the geographic area exceeding the threshold (from Figure 4). Similarly, only a small percentage of streams with reference condition characteristics (6 percent) or good benthic conditions (21 percent)

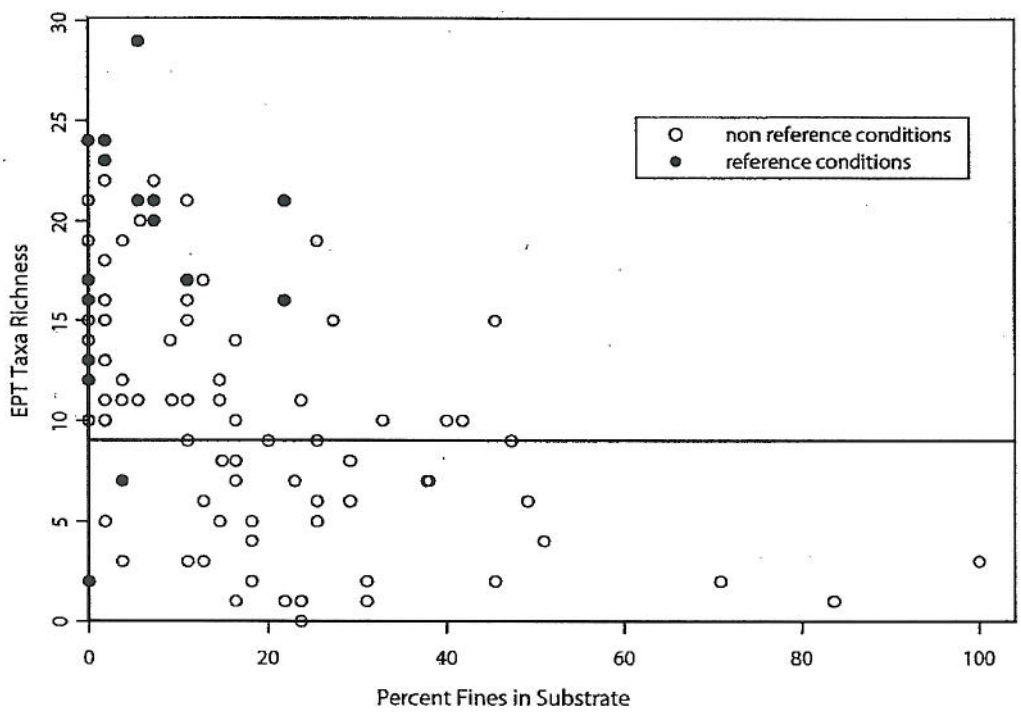


Figure 3. Plot of EPT Taxa Richness Against Percent Fines in Substrate (silt/clay fraction, less than 0.06 mm). Horizontal line for EPT taxa richness = 9. Solid circles are stream segments that exhibit reference condition characteristics. Open circles are segments not satisfying reference condition characteristics.

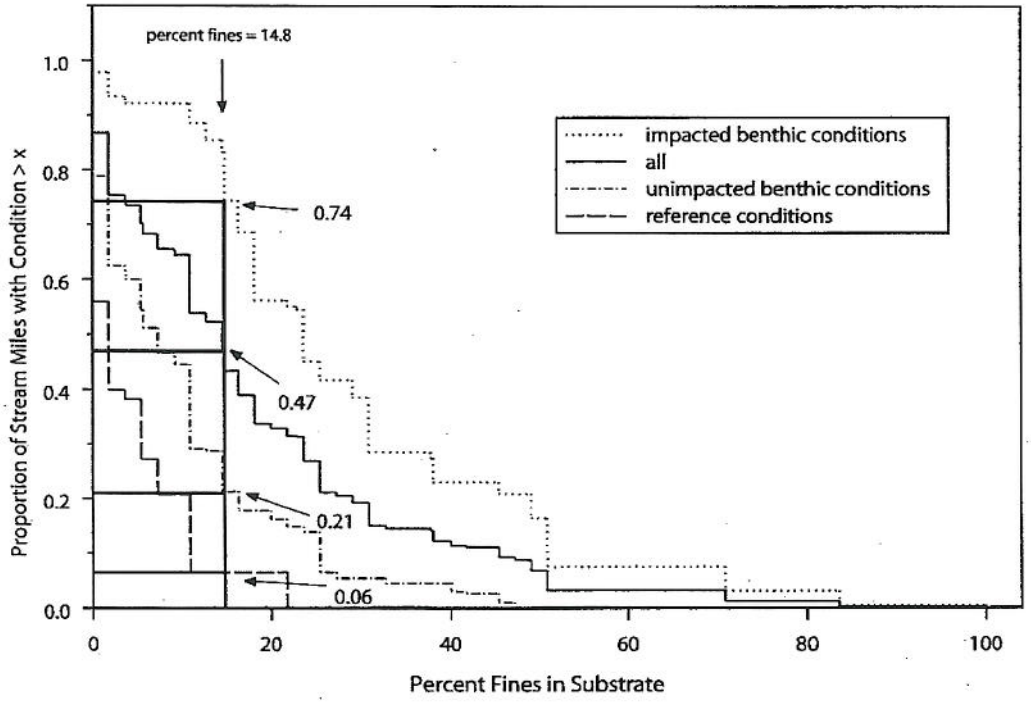


Figure 4. Reverse Cumulative Distribution Function (CDF) for Percent Fines in the Substrate (silt/clay fraction, less than 0.06 mm) for Stream Miles Across Entire Area, and Reverse Conditional CDFs of Stream Miles for Impacted Benthic Conditions (EPT taxa richness less than 9), Unimpacted Benthic Conditions (EPT taxa richness equal to or greater than 9), and Reference Conditions. Horizontal lines are where the threshold of 14.8 percent fines intersects curves.

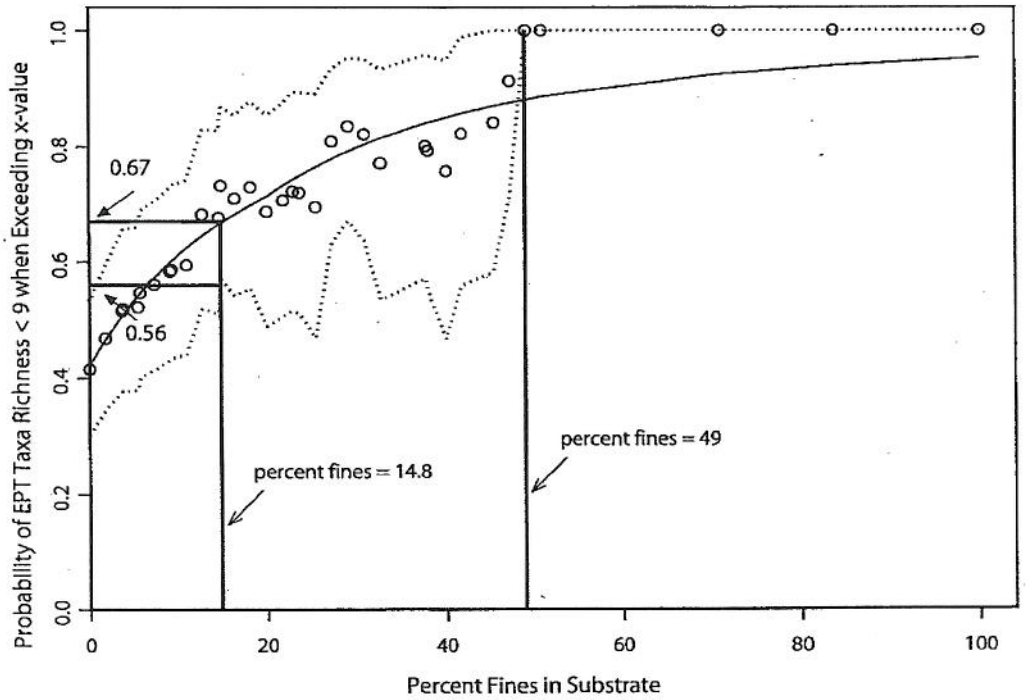


Figure 5. Probability of Observing EPT Taxa Richness Less Than 9 (benthic impact) in Mid-Atlantic Streams (open circles) if Specified Value of Percent Fines in the Substrate (silt/clay fraction, less than 0.06 mm) is Exceeded. Solid line is fit of Equation (4) (see Table 1). Dotted lines are 95 percent confidence intervals (CIs) from bootstrap estimation.

would exceed the 14.8 percent fines threshold, but a much larger percentage of impacted streams (74 percent) would exceed it (from Figure 4). These values provide an estimate of the number of "false positives" for this value of a threshold for percent fines as the indicator of sedimentation. Because multiple stressors often impact stream communities, one cannot estimate the "false negatives." A community not stressed by the stressor of interest might be stressed in some other way.

The coefficients from the nonlinear least squares regression for Equation (4) are given in Table 1, with the fitted curve shown in Figure 5. This technique also determined a threshold of impact of 14.8 percent fines in the substrate (Table 2). Using the nonparametric deviance reduction technique, a threshold of impact of 15.3 percent fines in the sediment, with p = 0.03, was identified. All three techniques for identifying a threshold of impact from the conditional probability analysis yielded consistent results (Table 2).

TABLE 1. Coefficients (mean value and confidence limits) From Nonlinear Least Squares Regression of Equation (4) for Percent Fines in Substrate Against Probability of Impacted Benthic Community (EPT taxa richness less than 9).

	Mean	90 Percent CI	95 Percent CI
B_0	-0.0328	(-0.049, -0.03)	(-0.0516, -0.0292)
B_1	-0.159	(-0.194, -0.138)	(-0.202, -0.133)
x_0	14.8	(13.4,15.9)	(13.1, 16.2)

These three different techniques all determined thresholds of impact that separated the data such that a difference could be detected from what would be expected for the entire geographic area. For the first technique, the existence of nonoverlapping CIs provided the difference. For the second technique, the nonoverlap of CIs for the curvature parameters established the difference. In the third technique, the null hypothesis of no changepoint was rejected (p = 0.03).

The literature supports a biological response of fishes to the thresholds for percent fines in the substrate: survival of salmonids have been shown to be negatively affected when fines exceed 10 to 20 percent (McNeil and Ahnell, 1964; Burns, 1970; Tappel and Bjornn, 1983; Chapman, 1988; Peterson *et al.* 1992; Argent and Flebbe, 1999). These studies for salmonids were all for a larger sediment size range (silt/clay/sand) than was chosen for purposes of this demonstration. However, the silt/clay fraction is always less than or equal to the silt/clay/sand fraction of the same sample. These threshold values are consistent with the lower end of reported response levels in the literature (see Newcombe and Jensen, 1996).

DISCUSSION

Scientifically defensible numeric criteria are highly desirable for water quality protection programs responsible for preventing the impairment of aquatic systems. Historically, for single chemical pollutants, carefully controlled laboratory bioassays have been conducted. From the dose-response relationship derived from these bioassays, an appropriate criterion for the pollutant was developed that is protective of aquatic life. Unfortunately, this historical approach is not applicable for nonpoint source pollution. At low levels, nonpoint source materials may not be a pollutant, but may be necessary for the functioning of the aquatic systems (e.g., nutrients, sediments) and their levels may naturally fluctuate over a geographic area. Only when the levels become excessive (usually in conjunction with anthropogenic activity), do they

TABLE 2. Summary of Thresholds of Impact for Percent Fines in Substract (silt/clay fraction, less than 0.06 mm) Identified by Conditional Probability Analysis (mean threshold and confidence intervals, when available) Using Three Techniques.

	Threshold (percent)	90 Percent CI (percent)	95 Percent CI (percent)
Nonoverlapping Confidence Intervals	12.7[1] 14.8[2]		
Change in Curvature of Fitted Curve	14.8	13.4 to 15.9	13.1 to 16.2
Nonparametric Deviance Reduction	15.3 (p = 0.03)	11.8 to 26.4	8.9 to 26.4

[1]From two-sided 90 percent CI.
[2]From two-sided 95 percent CI.

become pollutants and require criteria development for their control. The development of scientifically defensible approaches for establishing thresholds, and eventually criteria, for nonpoint source pollution is a critical need for regulatory agencies (USEPA, 2003b). Any approach undertaken to develop nonpoint source criteria must take into account the natural variability of the pollution occurring across the geographic area of interest, and the impact of other stressors that are likely impacting the aquatic systems as well.

Use of the CPA can establish realistic thresholds for the impact on stream biotic condition by nonpoint source pollution. This approach was applied to establish a threshold of sediment impact on a susceptible biological community in wadable streams in the mid-Atlantic region of the U.S. The mid-Atlantic was selected because of the extensive amount of research and monitoring of streams in this region (e.g., Boward *et al.*, 1999; USEPA, 2000b), which provided the information base needed that would satisfy the conditions for application of CPA. The necessary conditions were: (1) monitored data must be collected based on a probability based sampling design; (2) there must be some metric that can quantify the pollution parameter of interest; (3) there must be a response metric sufficiently sensitive to respond to the extant levels of the pollution parameter of interest; (4) independent studies must be available that identify the characteristics of an impacted response metric; and (5) the pollution parameter must be capable of exerting a strong effect on the response metric.

The streams in the mid-Atlantic region met these criteria. Sufficient empirical data were available from a probability monitoring design for wadable streams in the region (see McDonald *et al.*, 2004; data are available from http://www.epa.gov/emap/). The probabilistic sampling allows for statistically rigorous extrapolation from the sites sampled to the entire region of interest. Sedimentation was a major stressor in these streams (Boward *et al.*, 1999; USEPA, 2000b), and sufficient information was available on the percentage of fines in the substrate to allow its use as a surrogate for sedimentation (Klemm *et al.*, 2002). Data had been collected on EPT taxa richness at these sites and related to stream condition (USEPA, 2000b). Davis and Scott (2000) had determined a level of EPT taxa species richness (less than 9) below which mid-Atlantic highland streams were likely to be impacted and in relatively poor condition. Last, EPT taxa richness responded strongly to sedimentation (Figure 3).

It was decided not to develop thresholds for protecting against impact based on not exceeding a pollution metric value (i.e., $P(y = 0 \mid x \leq x_C)$, where $y = 0$ represents unimpacted conditions). The approach of looking for a threshold which, if not exceeded, would indicate a high probability of encountering unimpacted conditions, is the approach taken for criteria developed from laboratory toxicological studies. What makes this appropriate for laboratory studies is the ability to control for all stressors other than the one for which criteria are being developed. These studies provide $y = 0$ (unimpacted) if $x < x_C$. This approach does not work when dealing with actual field data, as one cannot control for all of the myriad stressors that are affecting the biological communities. The biological response observed reflects the cumulative response to all of the stressors. As the data for a specific stressor is analyzed, and as the magnitude of that stressor decreases, one would not expect a continual increase in the likelihood of unimpacted conditions, unless all of the stressors are strongly correlated. Reducing one stressor would still leave other stressors eliciting impact (in the case of 42 percent of the stream miles in Figure 2). Therefore, any calculation of $P(y = 0 \mid x \leq x_C)$ would likely be confounded by other stressors (i.e., value for unconditional probability would not be zero). Thus, the thresholds identified with the conditional probability analysis (and based on the reverse CCDF) are thresholds of impact, above which the likelihood of impact is high. One is able to pull out the signal in the mixture of multiple stressors with the conditional probability analysis because the stressor chosen was strong enough to elicit an impact as the magnitude of the stressor increased. With the conditional probability analysis approach, one is protecting the aquatic resource against the likelihood of impact.

The EPT taxa richness metric was used to identify impacted stream communities. The taxa in the orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) respond similarly to sedimentation as estimated by percent fines in the substrate (Figure 3), with the probability of impact increasing as percent fines increased. Other functional biological groupings that responded similarly to EPT were benthic invertebrate scrapers and intolerant taxa richness, while noninsect benthic invertebrates, benthic invertebrate scavengers, and tolerant taxa richness responded with the probability of impact decreasing as percent fines increased. Conditional probability plots could have been generated using any of these groupings of the benthic invertebrate community, if a level of biological impact could be independently assigned (similar to EPT taxa richness less than 9).

With the CPA approach, traditional statistics cannot be used to ascertain the threshold of impact. Instead, a weight-of-evidence approach based on three separate techniques was used. However, this assumed that there was a consistency in the threshold levels identified using these disparate techniques. The CPA

does provide relatively consistent thresholds of impact for the percent fines in the substrate, irrespective of which of the three techniques were applied (Table 2). These CPA thresholds contrasted markedly with the threshold values obtained with the two *ad hoc* approaches currently practiced for developing aquatic criteria based on monitoring data from sites across a geographic area (USEPA, 2000a). The two techniques consist of setting thresholds with either the levels of stressor associated with streams in the 75th percentile of the reference stream miles sampled or the 25th percentile of all stream miles sampled. Using these approaches, the threshold for percent fines would be approximately 1.9 percent based on all stream miles and approximately 7.1 percent based on reference stream miles (Figure 4). These values are substantially lower than estimates from this study, and fall outside of the 95 percent confidence limits (Table 2). These values are also substantially lower than literature thresholds for percent fines.

While the literature supports the biological importance associated with the thresholds identified, the agreement of these threshold values with lower values from the literature does not validate the thresholds. However, it does gives credence to the conditional probability analysis approach for identifying realistic thresholds for use in development of criteria for protection of aquatic life. The CPA can provide environmental managers with an additional tool to evaluate the tradeoffs of setting different criteria. Using CPA, environmental managers can examine a given criterion and the tradeoffs: the likely number of stream miles that actually have good biological communities when the criterion level is exceeded and the number of streams that have impacted biology when the criterion level is not exceeded. This would allow the protection of the ecosystems to be more quantitative and explicit when being weighed in conjunction with economic considerations.

The CPA approach for threshold of impact can be combined with information on reference conditions and toxicological data to develop candidate values for water quality criteria. The steps in this process that provide the candidate values are listed below.

1. Acquire the survey data (probability based) that includes candidate pollutant for criterion development and biological response metrics.
2. Use available information on reference conditions (physical, chemical, and habitat metrics) to define impacted biological conditions in terms of biological metrics.
3. Conduct conditional probability analysis (probability of impact if value of candidate pollutant is exceeded).
4. Identify threshold of impact from conditional probability analysis results.
5. Evaluate identified threshold of impact against reference conditions, impacted conditions, and good conditions. Evaluate identified threshold for biological importance.

Nonetheless, additional work must be done to evaluate and validate the conditional probability analysis approach for identification of thresholds of impact. This could be accomplished by using other survey data from other geographic areas where the conditions for CPA are met. The approach should also be tested with other pollution and response parameters to confirm that this is a robust approach, which can be used for identifying realistic thresholds of impact.

SUMMARY

The conditional probability analysis approach can be used to develop realistic thresholds of impact for nonpoint source pollution on aquatic benthic communities in waterbodies across a region. However, this approach is predicated on the following conditions: (1) monitored data have been collected based on a probability based sampling design; (2) some metric must be available that can quantify the pollution parameter of interest; (3) a response metric sufficiently sensitive to respond to the extant levels of the pollution parameter of interest must be available; (4) independent studies must be available that identify the characteristics of an impacted response metric; and (5) the pollution parameter must be capable of exerting a strong effect on the response metric. In the example presented here, realistic thresholds of impact on EPT taxa richness due to sedimentation in mid-Atlantic wadable streams were identified. Threshold values from CPA were found to be in the range of 12 to 15 percent fines in the substrate, based on three different techniques for threshold identification. These threshold values were found to be consistent with existing literature from laboratory and field studies. These values were quite different from those determined with the current *ad hoc* practice, with the current practice appearing to produce overly prescriptive thresholds. Development of scientifically defensible thresholds are a necessary first step for managers in establishing protective criterion. However, thresholds determined with CPA, or other methods, should not be used exclusively to set water quality criteria, as other additional factors (e.g., designated uses, ecotoxicological data, economics) must be considered by managers when establishing criteria and standards.

ACKNOWLEDGMENTS

Thanks to Brian Hill, Jim Wickham, Walter Berry, Jerry Pesch, John Van Sickle, Phil Kaufman, and anonymous reviewers for the critical and constructive reviews that they provided on various versions of this manuscript. Special thanks goes to Steve Hedtke for the encouragement to apply the conditional probability analysis approach to sediment criteria and to the EMAP-Surface Waters team for generating high quality data and making these data available to all. S-Plus software was used for statistical analyses and graphical displays. The research described in this paper was funded by the U.S. Environmental Protection Agency. This paper was not subjected to Agency review, and therefore does not necessarily reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

LITERATURE CITED

Argent, D.G. and P.A. Flebbe, 1999. Fine Sediment Effects on Brook Trout Eggs in Laboratory Streams. Fisheries Research 39(3): 253-262.

Austin, P.C. and J.E. Hux, 2002. A Brief Note on Overlapping Confidence Intervals. Journal of Vascular Surgery 36(1):194-195.

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, D.C.

Bash, J., C. Berman, and S. Bolton, 2001. Effects of Turbidity and Suspended Solids on Salmonids. University of Washington, Center for Streamside Studies. *Available at* http://depts.washington.edu/cwws/Outreach/Publications/Salmon%20and%20Turbidity.pdf. *Accessed in* August 2005.

Berry, W., N. Rubinstein, B. Melzian, and B. Hill, 2003. The Biological Effects of Suspended and Bedded Sediment (SABS) in Aquatic Systems: A Review. U.S. Environmental Protection Agency, Office of Research and Development, Internal Report. *Available at* http://www.epa.gov/waterscience/criteria/sediment/appendix1.pdf. *Accessed in* August 2005.

Best, E.P.H., C.P. Buzzelli, S.M. Bartell, R.L. Wetzel, W.A. Boyd, R.D. Doyle, and K.R. Campbell, 2001. Modeling Submersed Macrophyte Growth in Relation to Underwater Light Climate: Modeling Approaches and Application Potential. Hydrobiologia 444(1-3): 43-70.

Borja, A., V. Valencia, J. Franco, I. Muxika, J. Bald, M. J. Belzunce, and O. Solaun, 2004. The Water Framework Directive: Water Alone, or in Association With Sediment and Biota, in Determining Quality Standards? Marine Pollution Bulletin 49(1-2):8-11.

Boward, D.M., P.F. Kazyak, S.A. Stranko, M.K. Hurd, and T.P. Prochaska, 1999. From the Mountains to the Sea: The State of Maryland's Freshwater Streams. Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, EPA 903-R-99-023, Annapolis, Maryland.

Burns, J.W., 1970. Spawning Bed Sedimentation Studies in Northern California Streams. California Fish and Game Quarterly 56(4):253-270.

Canfield, D.E.G., K.A. Langeland, S.B. Linda, and T.W. Haller, 1985. Relations Between Water Transparency and Maximum Depth of Macrophyte Colonization in Lakes. Journal of Aquatic Plant Management 23:25-28.

Chapman, D.W., 1988. Critical Review of Variables Used to Define Effects of Fines in Redds of Large Salmonids. Transactions of the American Fisheries Society 117(1):1-21.

Cherry, S., 1996. A Comparison of Confidence Interval Methods for Habitat Use-Availability Studies. Journal of Wildlife Management 60(3):653-658.

Cherry, S., 1998. Statistical Tests in Publications of the Wildlife Society. Wildlife Society Bulletin 26(4):947-953.

Covich, A.P., 1999. The Role of Benthic Invertebrate Species in Freshwater Ecosystems. Bioscience 49(2):119-127.

Davis, W.R. and J. Scott, 2000. Mid-Atlantic Highlands Streams Assessment: Technical Support Document. EPA/903/B-00/004, Mid-Atlantic Integrated Assessment Program, Region 3, U.S. Environmental Protection Agency, Ft. Meade, Maryland.

Davis, W.S. and J.E. Lathrop, 1992. Freshwater Benthic Macroinvertebrate Community Structure and Function. *In:* Sediment Classification Methods Compendium, EPA 823-R-92-006, pp. 8-1 to 8-26. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Fisher, D.J. and D.T. Burton, 2003. Comparison of Two U.S. Environmental Protection Agency Species Sensitivity Distribution Methods for Calculating Ecological Risk Criteria. Human and Ecological Risk Assessment 9(3):675-690.

Gerritsen, J., J. Green, and R. Preston, 1994. Establishment of Regional Reference Conditions for Stream Biological Assessment and Watershed Management. *In:* Proceedings, Watershed '93, A National Conference on Watershed Management. EPA/840/R-90/002, U.S. Environmental Protection Agency, Office of Water, Washington, D.C., pp. 797-801.

Herlihy, A.T., D.P. Larsen, S.G. Paulsen, N.S. Urquhat, and B.J. Rosenbaum, 2000. Designing a Spatially Balanced, Randomized Site Selection Process for Regional Stream Surveys: The EMAP Mid-Atlantic Pilot Study. Environmental Monitoring and Assessment 63(1):95-113.

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

Hohreiter, D.W., and D.K. Rigg, 2001. Derivation of Ambient Water Quality Criteria for Formaldehyde. Chemosphere 45(4-5):471-486.

Insightful Corp., 2001. S-Plus 6 for Windows, Guide to Statistics, Volume 2. Insightful Corp., Seattle, Washington.

Jimenez, B., J. Ramos, and L. Quezada, 1999. Analysis of Water Quality Criteria in Mexico. Water Science and Technology 40(10):169-175.

Kaller, M.D., and K.J. Hartman, 2004. Evidence of a Threshold Level of Fine Sediment Accumulation for Altering Benthic Macroinvertebrate Communities. Hydrobiologia 518:95-104.

Kamizoulis, G. and L. Saliba., 2004. Development of Coastal Recreational Water Quality Standards in the Mediterranean. Environment International 30(6):841-854.

Kaufmann, P.R., P. Levine, E.G. Robinson, C. Seeliger, and D.V. Peck, 1999. Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003, U.S. Environmental Protection Agency, Washington, D.C.

Kaufmann, P.R. and E.G. Robinson, 1998. Physical Habitat Characterization in Environmental Monitoring and Assessment Program – Surface Waters. *In:* Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams, J.M. Lazorchak, D.J. Klemm and D.V. Peck (Editors). EPA/620/R-94/004F, U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C., pp. 77-118.

Kay, D., J. Bartram, A. Pruss, N. Ashbolt, M.D. Wyer, J.M. Fleisher, L. Fewtrell, A. Rogers, and G. Rees, 2004. Derivation of Numerical Values for the World Health Organization Guidelines for Recreational Waters. Water Research 38(5):1296-1304.

Klemm, D.J., K.A. Blocksom, W.T. Thoeny, F.A. Fulk, A.T. Herlihy, P.R. Kaufmann, and S.M. Cormier, 2002. Methods Development and Use of Macroinvertebrates as Indicators of Ecological Conditions for Streams in the Mid-Atlantic Highlands Region. Environmental Monitoring and Assessment 78(2):169-212.

Lazorchak, J.M., D.J. Klemm, and D.V. Peck (Editors), 1998. Environmental Monitoring and Assessment Program-Surface Waters. Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams, EPA/620/R-94/004F, U.S. Environmental Protection Agency, Washington, D.C.

Loch, D.D., J.L. West, and D.G. Perlmutter, 1996. The Effect of Trout Farm Effluent on the Taxa Richness of Benthic Macroinvertebrates. Aquaculture 147(1-2):37-55.

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

McDonald, M., R. Blair, D. Bolgrien, B. Brown, J. Dlugosz, S. Hale, S. Hedtke, D. Heggem, L. Jackson, K. Jones, B. Levinson, R. Linthurst, J. Messer, A. Olsen, J. Paul, S. Paulsen, J. Stoddard, K. Summers, and G. Veith, 2004. The Environmental Protection Agency's Environmental Monitoring and Assessment Program. *In:* Environmental Monitoring. G. B. Wiersma (Editor). CRC Press LLC, New York, New York, pp. 649-668.

McNeil, W.J. and W.H. Ahnell, 1964. Success of Pink Salmon Spawning Relative to Size of Spawning Bed Materials. U.S. Fish and Wildlife Service Special Scientific Report, Fisheries 469, Washington, D.C.

Newcombe, C.P. and J.O.T. Jensen, 1996. Channel Suspended Sediment and Fisheries: A Synthesis for Quantitative Assessment of Risk and Impact. North American Journal of Fisheries Management 16(4):693-727.

Niemi, G.L. and M.E. McDonald, 2004. Application of Ecological Indicators. Annual Review of Ecology, Evolution, and Systematics 35:89-111.

Perry, J. and E. Vanderklein, 1996.Water Quality: Management of a Natural Resource. Blackwell Science, Cambridge, Massachusetts.

Peterson, N.P., A. Hendry, and T.P. Quinn, 1992. Assessment of Cumulative Effects on Salmonid Habitat: Some Suggested Parameters and Target Conditions. University of Washington, Center for Streamside Studies, TFW-F3-92-001, Seattle, Washington.

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. U.S. EPA Report EPA/440/4-89/001, Washington, D.C.

Qian, S.S., R.S. King, and C.J. Richardson, 2003. Two Statistical Methods for the Detection of Environmental Thresholds. Ecological Modelling 166(1-2):87-97.

Rahlfs, V.W., 1997. Understanding and Evaluating Clinical Trials [Letter]. Journal of the American Academy of Dermatology 37(5):803-804.

Rausina, G.A., D.C.L. Wong, W. Raymon Arnold, E.R. Mancini, and A.E. Steen, 2002. Toxicity of Methyl-Tert-Butyl Ether to Marine Organisms: Ambient Water Quality Criteria Calculation. Chemosphere 47(5):525-534.

Russo, R.C., 2002. Development of Marine Water Quality Criteria for the USA. Marine Pollution Bulletin 45(1-12):84-91.

Spooner, J., L. Wyatt, S.W. Coffey, S.L. Brichford, J.A. Arnold, M.D. Smolen, G.D. Jennings, and J.A. Gale, 1991. Nonpoint Sources. Research Journal of the Water Pollution Control Federation 63(4):527-536.

Stevens, D.L., Jr., 1997. Variable Density Grid-Based Sampling Designs for Continuous Spatial Populations. Environmetrics 8:167-195.

Stevens, D.L., Jr. and A.R. Olsen, 1999. Spatially Restricted Surveys Over Time for Aquatic Resources. Journal of Agricultural, Biological, and Environmental Statistics 4(4):415-428.

Suttle, K.B., M.E. Power, J.M. Levine, and C. McNeely, 2004. How Fine Sediment in Riverbeds Impairs Growth and Survival of Juvenile Salmonids. Ecological Applications 14(4): 969-974.

Tappel, P.D. and T.C. Bjornn, 1983. A New Method of Relating Size of Spawning Gravel to Salmonid Embryo Survival. North American Journal of Fisheries Management 3:123-135.

USEPA (U.S. Environmental Protection Agency), 1994. Water Quality Standards Handbook (Second Edition). EPA/823/B-94/005, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

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

USEPA (U.S. Environmental Protection Agency), 2000b. Mid-Atlantic Highlands State of the Streams. EPA/903/R-00/005, U.S. Environmental Protection Agency, Region 3, Philadelphia, Pennsylvania.

USEPA (U.S. Environmental Protection Agency), 2002. National Water Quality Inventory: 2000 Report. EPA-841-R-02-001, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

USEPA (U.S. Environmental Protection Agency), 2003a. Developing Water Quality Criteria for Suspended and Bedded Sediments (SABS). Potential Approaches: A U.S. EPA Science Advisory Board Consultation. U.S. Environmental Protection Agency, Office of Water, Draft Report. *Available at* http://www.epa.gov/waterscience/criteria/sediment/sab-discussion-paper.pdf. *Accessed in* August 2005.

USEPA (U.S. Environmental Protection Agency), 2003b. Setting Priorities to Strengthen the Foundation for Protecting and Restoring the Nation's Waters, Strategy for Water Quality Standards and Criteria. EPA-823-R-03-010, Office of Science and Technology, Office of Water, U.S. Environmental Protection Agency, Washington, D.C.

USEPA (U.S. Environmental Protection Agency), 2004. Environmental Monitoring and Assessment Program (EMAP): Mid-Atlantic Streams 1993-96 Data Sets. *Available at* http://www.epa.gov/emap/html/dataI/surfwatr/data. *Accessed in* August 2005.

Venables, W.N. and B.D. Ripley, 1997. Modern Applied Statistics With S-Plus (Second Edition). Springer, New York, New York.

Waite, I.R., A.T. Herlihy, D.P. Larsen, and D.J. Klemm, 2000. Comparing Strengths of Geographic and Nongeographic Classifications of Stream Macroinvertebrates in the Mid-Atlantic Highlands, USA. Journal of the North American Benthological Society 19:429-441.

Yin, D., H. Jin, L. Yu, and S. Hu, 2003. Deriving Freshwater Quality Criteria for 2,4-Dichlorophenol for Protection of Aquatic Life in China. Environmental Pollution 122(2):217-222.

Young, M.K., W.A. Hubert, and T.A. Wesche, 1991. Selection of Measures of Substrate Composition to Estimate Survival to Emergence of Salmonids and to Detect Changes in Stream Substrates. North American Journal of Fisheries Management 11:339-346.

Zweig, L.D. and C.F. Rabeni, 2001. Biomonitoring for Deposited Sediment Using Benthic Invertebrates: A Test on 4 Missouri Streams. North American Benthological Society 20(4):643-657.

EXHIBIT 8

06/11/08

MEMORANDUM

From: William Swietlik, Chief
Ecological and Health Processes Branch
EPA/OW/HECD/OST

To: Robert Koroncai, Associate Director
Water Protection Division
EPA/Region 3

Subject: Development of Nutrient Endpoints for TMDLs in Pennsylvania

The Headquarters nutrient team and I have completed our review of the document entitled: *Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania*, prepared for Region 3 by Tetra Tech, Inc., dated November 20, 2007. It is our conclusion that the approach used in the document to derive the nutrient TMDL endpoints for use in implementing Pennsylvania's narrative standard is a scientifically defensible approach and is consistent with EPA guidance for deriving nutrient criteria.

The approach used in the document is an example of the multiple-lines-of-evidence (or weight-of-evidence) approach. The report examined different lines of evidence to derive nutrient numbers in three categories, and involved 17 different lines of evidence, constituting a very thorough analysis. The multiple-lines-of evidence approach is recommended by EPA in the following guidance.

- U.S. EPA. 2000b. *Nutrient Criteria Technical Guidance Manual: Rivers and Streams*. Office of Water, Office of Science and Technology, Washington, DC. EPA-822-B-00-002. In summary, this guidance states that a weight of evidence approaches that combines one or more of the three approaches; 1) Reference reaches, 2) Predictive relationships and, 3) Published threshold values; while considering downstream effects, will produce criteria of greater scientific validity.

- U.S. EPA. 2006. *Framework for Developing Suspended and Bedded Sediments (SABS) Water Quality Criteria.* Office of Water and Office of Research and Development, Washington, DC. EPA-822-R-06-001. This

document recommends an integration and synthesis of multiple methods. This recommendation is based on a conclusion of the USEPA Science Advisory Board that "no single method would suffice complete criteria development in every situation and that multiple methods applied simultaneously (synthesized) may be more appropriate for criteria development."

In October, 2007 EPA HQ provide training to the Region II and III States on the weight-of-evidence methodology and how it can be applied to developing numeric nutrient values. It is good to see the Region benefited from our training and that you are now employing this approach.

EXHIBIT 9

Department of Civil and Environmental Engineering
University of Delaware | Newark, DE 19716-3120
phone: 302-831-2442 | info@ce.udel.edu | fax: 302-831-3640

6/17/2008

John C. Hall, Esq.
Hall and Associates
1101 15th Street, NW
Washington, DC 20005

Re: Determining Appropriate Nutrient Reduction Requirements for Streams

Dear Mr. Hall:

This letter is in response to your recent inquiry regarding the establishment of nutrient standards to protect stream environments. I have devoted over 35 years of my professional career to the evaluation of nutrient impacts and other chemical interactions related to water quality and sediment criteria development. Nutrient standard development is a complex subject that is not amenable to simplified determinations. Stream environments, in particular, are subject to a variety of physical conditions that alter whether and how nutrients may stimulate excessive plant growth and significantly alter stream ecology. Any defensible analysis would have to account for these factors. This is a widely held view that is reflected in numerous articles in the peer reviewed literature and federal guidance documents.

You asked in particular for my views on a recent federal approach suggested for use in deriving nutrient standards for streams in Pennsylvania. As I understand it, the approach seeks to directly correlate total phosphorus (TP) "growing season" average concentrations with monitored invertebrate populations. It proposes to use a procedure known as "conditional probability" to identify an instream TP target that would protect invertebrate populations. You provided me with one such report that includes such an analysis. You asked, in general, whether the scientific literature would support the using such a direct correlation for development of a protective instream standard. The following provides a brief response

1. **Is it an accepted principle in the scientific literature that total phosphorus directly impacts invertebrate levels in streams?**

Answer: No. Nutrients do not act like toxic chemicals, e.g. copper. They do not have a direct impact on sensitive invertebrates. Therefore, it is not appropriate to directly compare nutrient levels in general and TP in particular to invertebrate responses. A scientifically defensible analysis must show how nutrients are affecting plant growth and then, how such plant growth is adversely impacting the ecology. It is well recognized in the literature that nutrients may cause an adverse impact to lake ecology if excessive algal productivity occurs. The relationships between nutrient loadings (*not* nutrient concentrations) and algal productivity in lakes can be

analyzed using a variety of simple and more complex models. The mechanisms by which excessive algae cause problems include the effects of low dissolved oxygen, reduced transparency and loss of fish habitat. The analogous relationships for streams are less well understood and documented. To document invertebrate impairments in streams that are related to excessive nutrients, it is first necessary to determine the level of nutrients that is causing excessive plant growth in that stream and then to relate the consequences of excessive plant growth, e.g. low dissolved oxygen, to its effect on stream biota.

2. **Does a correlation between total phosphorus concentrations and invertebrate levels demonstrate that the nutrient was the cause of the changing invertebrate level?**

Answer: No. First, correlation does not demonstrate causation and therefore, unless more is known (i.e., the elevated phosphorus is also documented to cause excessive plant growth) such analysis gives a preliminary indication, at best. Use of a TP concentration indicator for stream environments, in particular, is subject to confounding factors. Elevated total phosphorus may occur due to high erosion rates that are inimical to sensitive invertebrate because they destroy the habitat – a common problem encountered in smaller streams. Simply plotting total phosphorus versus invertebrate population could easily be misleading and result in regulating the wrong water quality parameter and the wrong source of impairment. Second, the use of total phosphorus is not the best indicator of the form of phosphorus that could stimulate excessive fixed plant growth in streams. It is widely understood that dissolved phosphorus, not total phosphorus, is the form that is used by plants. Therefore, such an analysis cannot prove that phosphorus is causing the identified change in invertebrate populations.

3. **What is the accepted scientific approach to show whether or not nutrients could be causing an impact on invertebrates in a stream?**

Answer: Develop a model that accounts for the relevant physical, chemical and biological factors influencing how a stream responds to nutrient inputs. Then relate the effects of increasing plant growth to the response of organisms expected to exist in the habitat in question. Such an evaluation could utilize a correlation type of assessment but only if the relevant physical and chemical factors were accounted for in the evaluation and only for waters with closely similar ecology were being reviewed. One important factor to consider is stream orientation, shading and water transparency as it is widely understood that the amount of incident light influences the degree of primary production occurring, regardless of the amount of nutrients present. The physical habitat (cobble, sandy, or rocky bottom) will also influence the types of sensitive invertebrates that may be present at a location, independent of water quality. Finally stream flow itself is an important factor.

In closing I would like to address one further issue, the appropriate application of conditional probability methodologies for setting criteria and standards. The conditional probability approach suggested for use in Pennsylvania has, to my knowledge, never been used to derive a federal numeric water quality objective. The problem is whether the variable in question, total phosphorus concentration, is directly related to the effect that is being evaluated. Or to put it

another way, is it clear that reducing the total phosphorus concentration would reduce the adverse effect, i.e. loss of benthic biota. This issue arose during the Science Advisory Board (SAB) review of the sediment criteria for which I was the technical director on behalf of EPA. We rejected this approach in favor of a method that directly links cause to effect.

One way to judge whether the conditional probability method is useful is to examine the probably plot itself. Note that most of the probability of an adverse effect ranges from 40% at the smallest TP concentration (5 ug/L) to 70% for approximately 100 ug/L.. Therefore large changes in concentration would be predicted to have only a small change in adverse effect probability: 70% to 40% for a twenty fold concentration change. It is likely that such a small change in adverse effect probability would not be detectable. Furthermore there are data used in the probability plot where *increasing* the total phosphorus causes a *decrease* in adverse effect probability. The reason for these is that there is only a weak relationship between TP concentration and adverse effects. Many other factors influence the benthic biota.

Therefore, while conditional probability can be considered as a possible tool for identifying streams for further investigation, it was not a sufficient basis to establish standards. In my opinion, such an approach cannot provide information on the pollutant level and form (e.g. dissolved or total phosphorus) that must be regulated to protect the environment.

One final comment: using this statistical procedure would be contrary to all the EPA criteria development that has preceded this effort. The scientific basis for the EPA water quality criteria is one of the landmark achievements of the agency. The methodology has been adopted almost universally. It has been reviewed many times. It is based squarely on the causal relationships between the chemical being regulated and the effects being protected. The proposed nutrient criteria would be a retreat from scientifically defensible criteria to simply an expedient solution for which little or no support exists. The claim that EPA employs the "best science" would not be true in this case. At the minimum I would strongly urge that an EPA Science Advisory Board or a National Academy of Science peer review be conducted to provide an independent evaluation of the proposed methodology.

Please let me know if I can provide any further information.

Sincerely,

Dominic M. Di Toro
Edward C. Davis Professor of Civil and Environmental Engineering
University of Delaware
Member of the National Academy of Engineering

EXHIBIT 10

TUFTS UNIVERSITY
School of Engineering

Professor and Louis Berger Chair in Computing and Engineering

July 8, 2008

John C. Hall, Esq.
Hall and Associates
1101 15th Street, NW
Washington, DC 20005

Dear Mr. Hall:

This letter is in response to your inquiry regarding the determination of appropriate nutrient reduction requirements for streams. Since 1974 when I first began studying the impact of phosphorus on Great Lakes water quality, I have devoted a great part of my scholarship to eutrophication. Although I initially focused on lakes and reservoirs, I shifted my focus to streams when I joined the faculty at the University of Colorado in 1986. It was there that I first recognized the significant and sometimes subtle differences between lakes and flowing systems like rivers and streams. Among other things, this resulted in my developing EPA-sponsored software that is expressly designed to simulate stream eutrophication (Chapra et al 2008).

Note that I have carefully read Prof. Di Toro's critique and wholeheartedly concur with all his conclusions. In particular, I strongly support his observations regarding the complexity of linking total P and invertebrate levels in streams. I also concur with his more general contention that the accepted scientific approach involves using models to quantify how streams respond to nutrient loadings. In the following, I will therefore limit my remarks to concerns that supplement his comments.

As I understand it, the proposed approach seeks to directly correlate total phosphorus (TP) concentrations with the health of invertebrate populations. Thus, phosphorus is treated as if it were a toxic substance that directly interferes with the viability and functioning of the biota.

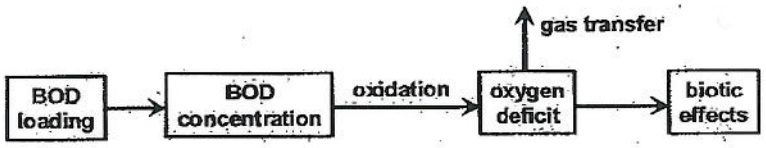
Department of Civil and Environmental Engineering
223 Anderson Hall
Medford, Massachusetts 02155
617 627-3654
Fax: 617 627-3994
Email: steven.chapra@tufts.edu

This is such a scientifically indefensible representation of the connection between nutrients and ecosystem health that I believe that its adoption would represent a grave mistake. Beyond being vulnerable to legal challenge, I am much more concerned that its adoption would ultimately be ineffective. That is, it could lead to costly controls that would not protect our precious stream ecosystems.

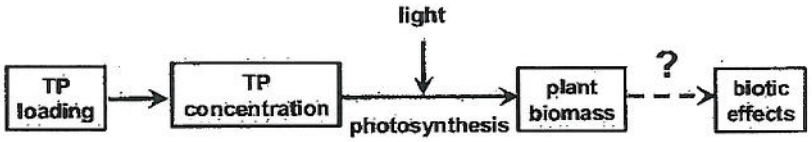
An analogy with one of the earliest water-quality management problems is instructive in illustrating my concern. In the early 20th century, point discharges of untreated urban sewage resulted in low oxygen concentrations in many of our nation's rivers and estuaries. Because adequate oxygen levels are necessary for most forms of aquatic life, the stress on ecosystems was great. Aside from direct impacts on fish and macroinvertebrates, low oxygen also triggers a number of secondary effects such as the generation of noxious odors.

Water-quality engineers recognized early on that the root cause of oxygen depletion was the presence of organic carbon and reduced nitrogen compounds in the urban wastewater. In the simplest sense, these compounds were collectively called biochemical oxygen demand or BOD. Once introduced into the river, the BOD was broken down by bacteria. Because the bacteria consumed oxygen to affect the break down, the river's oxygen resources were depleted.

As depicted in Figure 1*a*, the cause-effect sequence for this problem is BOD loading → BOD concentration → dissolved oxygen concentration deficit[1] → biotic impact. Notice that along with BOD oxidation, the deficit is also determined by gas transfer or reaeration.



(*a*) The BOD/Dissolved Oxygen Problem



(*b*) The TP/Biomass (AKA Eutrophication) Problem

Figure 1 Analogy between two river water-quality problems: (a) dissolved oxygen and (b) eutrophication. The question mark on the link between plant biomass and biotic effects is meant to suggest the complexity of this connection.

Although this is a very simplified representation of the problem, the important point is that the direct cause of the ecosystem impairment is the oxygen deficit, not the BOD concentration. Hence, although we might ultimately alleviate the ecosystem stress by reducing the BOD loading, the actual BOD concentration would in fact not be directly correlated with the low oxygen levels.

---

[1] The deficit measures the difference between the saturation concentration and the actual DO level (Figure 2). A high deficit, therefore, indicates low oxygen and a high ecosystem stress.

This can be clearly seen by using the classic Streeter-Phelps model to develop a plot of BOD and oxygen in a river below a single point source of BOD. Although this model is simple, it accurately captures the interplay between BOD and oxygen in one-dimensional rivers[2].

As in Figure 2, the BOD is highest at the discharge point ($t = 0$) where the BOD loading is introduced. It then decreases downstream as the BOD is broken down via oxidation. The BOD oxidation in turn consumes oxygen which leads to a rapid decrease in oxygen concentration. However, as the level plunges, oxygen transfer across the air-water interface increases. Eventually the oxygen profile levels off when the oxidation loss is balanced by reaeration. Thereafter, oxygen levels begin to climb as reaeration becomes dominant and the stream recovers.

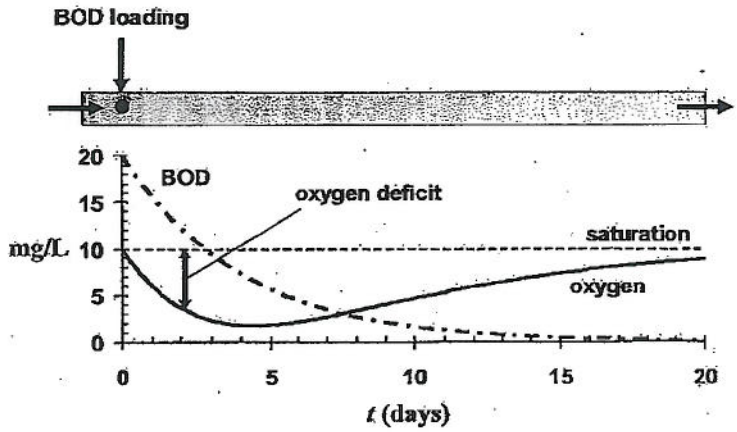


Figure 2 Simulation of BOD and dissolved oxygen versus travel time below a single point source into a one-dimensional river.

The most important feature of Figure 2 is that there is absolutely no spatial correlation between the in-stream BOD and oxygen levels. For example the highest BOD concentration and the lowest deficit both occur at the mixing zone ($t = 0$). Although BOD certainly causes oxygen depletion, it would therefore be ludicrous to specify an instream BOD concentration criterion in order to ensure an adequate oxygen level. Instead, the correct approach is to set an oxygen criterion that is directly connected with ecosystem health. A model can then be employed to link the oxygen concentration back to the BOD loading. This, of course, is how oxygen has been so effectively managed over the past century.

Beyond illustrating how river BOD and oxygen concentrations evolve and interact, employing a cause-effect model yields additional benefits. For example, rather than reducing the BOD load, the analysis suggests an alternative means to raise oxygen levels might involve enhancing gas transfer (e.g., oxygen diffusers, artificial waterfalls, etc.). Thus, the scientific approach reveals possible alternative remediation strategies that in certain cases might actually be more cost effective than source controls.

---

[2] In this context, one-dimensional means that changes only occur longitudinally and that the stream is well-mixed laterally (bank-to-bank) and vertically (with depth).

So how is this example relevant to the stream eutrophication problem? As illustrated in Figure 1*b*, a very simple representation of stream eutrophication is that nutrient loading (in this example, phosphorus[3]) results in increased stream nutrient concentration which in turn leads to increased plant biomass. Thus, as BOD is to oxygen deficit, phosphorus is to plant biomass. And just as it would be ludicrous to specify an instream BOD concentration criterion to solve the oxygen problem, it is equally misguided to specify an instream total phosphorus concentration criterion to solve the eutrophication problem.

Further, as described by Prof. Di Toro, the subsequent connection between increased biomass and biotic impacts is not as well understood as for oxygen. In fact, one such connection involves the impact of excessive plant growth on stream oxygen via both direct (photosynthetic gains and respiration losses) and indirect (when plants die they become BOD) pathways. However, the deleterious effects undoubtedly involve other non-oxygen related factors such as direct habitat impairment and shifts to undesirable plant species.

As with the previous oxygen example, beyond illustrating how river nutrients and biomass evolve and interact, the cause-effect model can point to alternative remediation approaches. For example, rather than reducing the nutrient load, Figure 1*b* suggests that an alternative means to reduce biomass might involve decreasing solar radiation. For example, this could be accomplished by planting riparian vegetation to create a canopy over the stream. As with oxygen, there might be cases where such alternatives could prove useful.

As a final technical note, it is important to distinguish between floating (i.e., phytoplankton) and attached (i.e., periphyton, filamentous algae and macrophytes) plants when dealing with stream eutrophication. For phytoplankton, which tend to dominate in deeper rivers, the direct analogy expressed in Figure 1 holds (see App. 1 for a detailed analysis). That is, it is absurd to set a TP criterion in order to manage river phytoplankton biomass.

For attached plants, the situation is much more complex. For systems dominated by macrophytes, nutrient management must consider whether the plants can draw nutrients from the sediments via their roots. For such situations, managing instream TP concentration would be counterproductive as photosynthesis would be effectively independent of water nutrients.

For filamentous algae (e.g., Cladophora) which draw nutrients directly from the water, there is certainly a closer connection. However, it is well known that (a) the photosynthesis rate of such organisms depends on their internal nutrient levels and (b) they only take up dissolved inorganic nutrients from the water. Hence, regulation based on water TP concentration would be ill-founded.

The same point can be made for periphyton but with a additional nuances. Whereas filamentous algae extend up into the water column, periphyton grow as biofilms on substrates such as bottom rocks. In such cases, their nutrient uptake can be influenced by transport limitations on the delivery of dissolved inorganic nutrients from the water into the biofilm. Such limitation would be dependent on stream hydraulics and hence related to factors such as stream velocity, depth, etc. Further, because they are bottom dwellers, their growth would obviously be highly dependent on the delivery of light to the stream bottom. Hence, along with nutrients, the clarity and depth of the overlying water would have to be considered in determining their

---

[3] Notice that in Figure 1, we have assumed that phosphorus is the limiting nutrient. It should be understood that this is an assumption and that another nutrient (e.g., nitrogen) or light might in fact be limiting.

biomass. Consequently, a simple, direct correlation with water TP concentration would seem to be overly simplistic.

In conclusion, I hope that the foregoing provides some indication of my great concern over the issue of stream eutrophication management. As a concerned environmentalist, as well as a lifelong fly fisherman, I truly support effective regulations to protect our nations great rivers and streams. However, as an environmental scientist and engineer, I also know that without a sound scientific basis, such regulations are likely to fail. It is in this spirit that I strongly support Prof. Di Toro's suggestion that at the minimum, an EPA Science Advisory Board or a National Academy of Science peer review be conducted to provide an independent evaluation of the proposed nutrient criteria.

Please let me know if I can provide any further information.

Sincerely,

Steven C. Chapra, Ph.D.

APPENDIX 1.
Why TP Concentration Standards are Inappropriate for Managing Phytoplankton Biomass in Rivers

This appendix attempts to address the question of why anyone would ever suggest that a total phosphorus criterion would represent a sensible strategy for managing river and stream eutrophication dominated by phytoplankton. In brief, I believe that the idea of river total phosphorus criteria originates from the misguided notion that effective lake management approaches can be seamlessly (and thoughtlessly) transferred to rivers and streams.

In the late 1960's and early 1970's, several limnologists suggested that total phosphorus concentration could serve as an effective trophic state indicator. In particular, Richard Vollenweider posited that lakes with total phosphorus concentrations less than 10 μgP/L would tend to be oligotrophic whereas those with greater than 20 μgP/L would tend be eutrophic.

Although Vollenweider himself repeatedly stated that these were approximate guidelines and not hard thresholds, the values were adopted by many lake managers as quantitative goals for managing lake eutrophication. And in fact, the approach has been a useful component of nutrient remediation schemes for a number of important systems including the Laurentian Great Lakes.

So why might the approach work for lakes and not for streams? The answers to this question lies in fundamental differences between these two types of natural waters.

In effect, the viability of the Vollenweider approach is predicated on the functioning of the particular lakes he studied. In particular, the approach was developed for deep, stratified, phosphorus-limited, North-temperate, lakes with long residence times (i.e., greater than a year). In such lakes, Vollenweider (and others) assumed that the spring total phosphorus concentration was a prime controlling factor of plant production over the ensuing summer growing season.

For this assumption to strictly hold, once the lake stratifies in late spring, the epilimnion must essentially behave as a batch or closed system. Thus, plant growth over the ensuing summer is primarily dictated by the finite store of nutrient represented by the spring phosphorus concentration rather than by external loads. The average summer level of biomass is then determined by the recycle of this pool between inorganic and organic forms. Empirical support for the approach was provided by a number of empirical correlations. The chief examples of these were logarithmic plots suggesting strong correlations between summer average chlorophyll *a* concentrations and spring total phosphorus concentration.

A simple computation can be used to illustrate how such an approach breaks down in rivers and streams. First, total phosphorus can be divided into three components

$$TP = p_p + p_i + p_o \tag{1}$$

where p_p = phytoplankton phosphorus (μgP/L), p_i = inorganic phosphorus (μgP/L), and p_o = non-phytoplankton organic phosphorus (μgP/L). If the chlorophyll *a* to phosphorus ratio is assumed to be 1 μgA/μgP, this means that p_p can be directly interpreted as a measure of phytoplankton biomass.

The river can be idealized as a steady-state, plug-flow system with a single point source of phosphorus (Figure 3). Further it is assumed that the river has uniform, steady flow and constant hydrogeometric properties (i.e., depth, width, etc.). For such cases, velocity will be constant and travel time and distance are linearly related (i.e., distance = velocity times travel time). Under these conditions, the following mass-balances can be written for each phosphorus component as

$$\frac{dp_p}{dt} = k_g \frac{p_i}{k_{sp} + p_i} p_p - k_r p_p - k_d p_p - k_s p_p \tag{2}$$

$$\frac{dp_i}{dt} = -k_g \frac{p_i}{k_{sp} + p_i} p_p + k_r p_p + k_h p_o \tag{3}$$

$$\frac{dp_o}{dt} = k_d p_p - k_h p_o \tag{4}$$

where t = travel time (d), k_g = maximum growth rate at constant light and temperature (/d), k_{sp} = phosphorus half-saturation constant (μgP/L), k_r = respiration/excretion rate (/d), k_d = death rate (/d), k_s = settling rate (/d), and k_h = hydrolysis rate (/d).

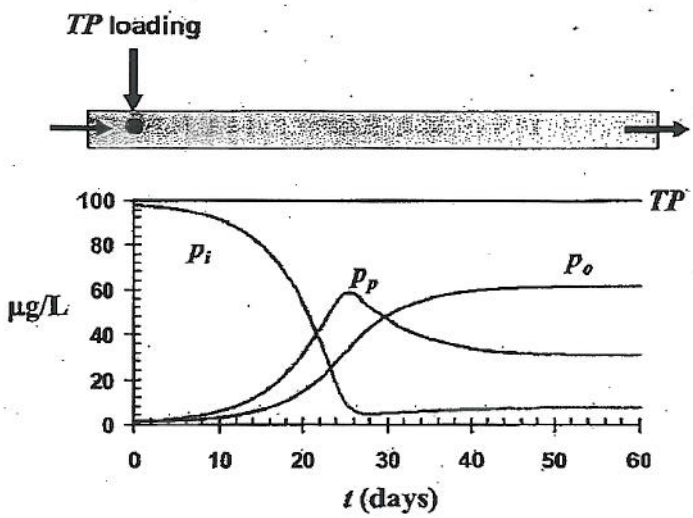


Figure 3 Simulation of phytoplankton, inorganic and organic phosphorus downstream from a point source.

Given reasonable values for the parameters and a set of initial conditions at the mixing point (Table 1), these equations can be integrated numerically to simulate how the various phosphorus species change as the water travels downstream. For the present example, the initial conditions are set so that the river has a high level of available, inorganic nutrient at the mixing point as would be the case for a high phosphorus discharge into an effluent-dominated river. In addition, the phytoplankton settling velocity is set to zero.

Table 1 Parameters and initial conditions used to simulate phytoplankton and phosphorus concentrations below a single point source to a one-dimensional river.

Parameter	Value	Units
k_g	0.5	/d
k_{sp}	5	µgP/L
k_r	0.2	/d
k_d	0.1	/d
k_s	0	/d
k_h	0.05	/d
Initial conditions:		
p_p	1	µgP/L
p_i	98	µgP/L
p_o	1	µgP/L

As displayed in Figure 3, because the inorganic P concentration is well above the half-saturation constant, the phytoplankton initially grow rapidly as the inorganic phosphorus is efficiently converted to phytoplankton biomass. Growth continues until the inorganic phosphorus level approaches the half saturation constant whereupon a peak is reached. At this point, growth has become sufficiently limited that it is exactly balanced by the respiration and death losses. Thereafter, the phytoplankton levels decline until the solution approaches a stable steady state. This asymptote represents the point at which phytoplankton growth exactly balances phosphorus recycle.

Note that because of the assumption of zero settling, the total P concentration is constant. This allows the component concentrations at the stable steady state to be computed exactly as

$$p_i = \frac{k_r + k_d}{k_g - (k_r + k_d)} k_{sp} \tag{5}$$

$$p_o = \left(1 - \frac{k_h}{k_d + k_h}\right)(TP - p_i) \tag{6}$$

$$p_p = \frac{k_h}{k_d + k_h}(TP - p_i) \tag{7}$$

Thus, we see that the ultimate inorganic phosphorus concentration is equal to the half saturation constant multiplied by the ratio of the phytoplankton loss rates ($k_r + k_d$) to the maximum net phytoplankton growth rate ($k_g - k_r - k_d$). The organic P and phytoplankton P concentrations are then dictated by the product of the total organic P (i.e., organic P and phytoplankton P) and a dimensionless number quantifying the relative values of the hydrolysis and death rates.

Although this is a very simple model, it dramatically illustrates why specifying a phosphorus concentration standard for rivers is ill-founded. Notice that until the asymptote is approached, there is no direct correlation between phytoplankton biomass and the total phosphorus concentration (as well as with any of the individual phosphorus species).

Just as is the case for BOD and oxygen, although phosphorus certainly causes increased phytoplankton biomass, there is absolutely no direct spatial correlation between in-stream TP and biomass. Hence, whereas a phosphorus standard makes some sense for a long residence time lake, it falls apart for a plug-flow system like a river.

Congress of the United States
Washington, DC 20515

November 7, 2008

Mr. Stephen L. Johnson
Administrator
United States Environmental Protection Agency
1200 Pennsylvania Ave NW
Washington, DC 20460

Dear Administrator Johnson:

We are concerned about the Region III in-stream standards for nutrients that the Environmental Protection Agency (EPA) has proposed to create new Total Maximum Daily Loads (TMDLs) for certain streams and creeks in Pennsylvania. Specifically, we are worried that the science used to develop the new discharge standards has not been peer-reviewed under the normal EPA protocol. Considering the practical implications, we could be imposing billions of dollars in compliance costs on Pennsylvania towns to meet standards that may or may not be necessary or achievable.

There is legitimate debate among nationally recognized scientists both for and against the strict new standards proposed by EPA. In fact, we know of two experts whom EPA has worked with in the past, Dr. Dominic DiToro and Dr. Stephen Chapra, who have said that the new standards are neither scientifically sound nor likely to restore streams. Given such strong statements, we find it odd that EPA would consider adopting these standards without a thorough vetting in a peer review process. Therefore, we strongly urge you to initiate a formal scientific peer review process and delay implementing new TMDLs until this process can be completed.

We agree that protecting and restoring Pennsylvania's streams and water bodies is critically important. But we also know that Pennsylvania communities cannot be expected to shoulder an expensive burden when it may not even achieve the intended restoration goals. At a time when we are putting so much work into strengthening Pennsylvanian's economy and creating jobs, we are reluctant to risk that progress when we are not assured that we are meeting restoration goals in the most cost-effective manner possible.

PRINTED ON RECYCLED PAPER

Thank you in advance for your attention to this important matter. We look forward to your response regarding formal peer review of the science for the proposed TMDLs and our continued work to protect the precious resources of Pennsylvania.

Sincerely,

Robert P. Casey, Jr.
United States Senator

Arlen Specter
United States Senator

Tim Holden
Member of Congress

Charles W. Dent
Member of Congress

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
WATER

DEC 24 2008

John C. Hall
Hall & Associates
1101 15th Street, N.W.
Suite 203
Washington, D.C. 20005

Dear Mr. Hall:

I am writing in further response to your letter, dated August 21, 2008, requesting peer review of the approach U.S. Environmental Protection Agency (EPA) Region III used to develop numeric endpoints for nutrients as part of the establishment of certain total maximum daily loads (TMDLs) in Pennsylvania. I am also following up on our meeting with you and your clients on November 13, 2008.

During the past few months, we have given careful consideration to the technical and legal issues you have raised about Region III TMDLs and, in particular, to your request that EPA conduct a peer review of the applicability of conditional probability to nutrient endpoint development. As you know, conditional probability was used in four of the seventeen lines of evidence the Region considered in determining an appropriately protective nutrient endpoint for the Pennsylvania TMDLs. Following our review, we continue to believe that these TMDLs, including the development of the nutrient endpoints using conditional probability-based and other lines of evidence, are technically and legally sound. In addition, we believe the implementation strategies accompanying the TMDLs provide Pennsylvania permit writers and permittees with sufficient flexibility to set and implement effluent limits that will ultimately achieve the necessary nutrient reductions in these watersheds.

Based on our review, we have also concluded that other States may wish to consider this approach and may benefit from more general guidance on its application. Therefore, we have decided to peer review application of this approach specifically to nutrients in the context of providing broader national guidance. At this time, we anticipate seeking peer review of a technical memorandum describing the method and how it could be used nationally to develop nutrient endpoints and criteria. Following successful completion of the peer review, we intend to update the technical memorandum as appropriate and make it available as a supplement to EPA's nutrient guidance. We

Printed on Recycled Paper

anticipate it will take 4-6 months to develop the technical memorandum and complete the peer review.

If you have any questions, or wish to discuss this further, please contact me, or call Jim Curtin in our Office of General Counsel at (202) 564-5482.

Sincerely,

Ephraim King, Director
Office of Science and Technology

cc: Jon Capacasa
Robert Koroncai
Jim Curtin

BNA, INC.

ENVIRONMENT REPORTER

BNA Employee-Owned Since 1947

Reproduced with permission from Environment Reporter, 40 ER 1593, 07/03/2009. Copyright © 2009 by The Bureau of National Affairs, Inc. (800-372-1033) http://www.bna.com

CLEAN WATER ACT

NUTRIENT STANDARDS

The authors of this article explore the Environmental Protection Agency's historical approach to nutrient regulation in streams and explain why they believe prior efforts were unsuccessful. They also provide a critique of EPA's latest proposal and say they hope a more effective and comprehensive approach to stream restoration can be established. The authors say the federal government's single-minded focus on nutrient discharges needs to be rethought if real progress is to be attained. In many circumstances, they suggest, alternative approaches such as bank and canopy restoration and other non-point source control efforts would produce far greater ecological benefits at far lower costs.

Critical Evaluation of EPA Stream Nutrient Standard Initiatives

By John C. Hall and William T. Hall

Since promulgation of the Clean Water Act in the 1970s and implementation of technology-based treatment requirements, such as secondary treatment for publicly owned treatment works, the nation's water quality has improved dramatically. Current water quality concerns are rarely attributed to excessive point source loads of oxygen-demanding pollutants or toxics. Rather, the largely uncontrolled discharge of nutrients from point and non-point sources continues to contribute to designated use impairments due to the anthropogenic eutrophication of lakes and streams, characterized by excessive plant growth. In a 1996 report to Congress, the Environmental Protection Agency estimated that forty percent of the rivers were impaired due to nutrient enrichment, fifty-one percent of the surveyed lakes, and fifty-seven percent of the surveyed estuaries were similarly adversely affected. Nutrients have also been implicated with the large hypoxic-zone in the Gulf of Mexico and hypoxia observed in several coastal waters (e.g., Long Island Sound, Chesapeake Bay).[1]

[1] Fact Sheet, *National Strategy for the Development of Regional Nutrient Criteria* (June 1998), available at http://www.epa.gov/waterscience/criteria/nutrient/strategy/nutsi.html.

Historically, eutrophication issues have been addressed on a watershed basis through total maximum daily loads (TMDLs) targeting compliance with dissolved oxygen (DO) standards. In freshwater streams, these TMDLs typically target reductions in phosphorus to limit plant growth such that diurnal DO swings do not result in an excursion of the minimum DO standard. However, plant growth dynamics in smaller streams are very complicated and good predictive relationships are not available. Numerous factors other than nutrient levels may control the degree of plant growth occurring in streams. As a consequence, this approach has not worked to define appropriate nutrient water quality criteria for flowing waters.

In response to criticisms that state's have been unable to define reasonable relationships between excessive plant growth and nutrient loads, EPA has abandoned this primary objective of nutrient control. EPA is now recommending that states use a ''weight-of-evidence'' approach to develop nutrient standards. This new approach to stream nutrient criteria development represents a very disturbing change in EPA's philosophy toward criteria development. The ''cause and effect'' basis underlying all other water quality standards has been abandoned in favor of an approach that has no specific linkage to use impairment. As a result, the ability of the resulting endpoints to mitigate nutrient-related impairments is even more uncertain. Application of the weight-of-evidence approach has resulted in extremely restrictive nutrient limitations unrelated to site-specific needs. In response to concerns raised by affected parties on several TMDL actions, EPA has agreed to a full Science Advisory Board (SAB) peer review of the agency's new approach and has requested nominations of experts for this review (74 *Fed. Reg.* 19, 085, 4/27/09). This review is expected to take place later this year.

The authors believe that the federal government's single-minded focus on nutrient discharges needs to be rethought if real progress is to be attained. In many circumstances, alternative approaches such as bank and canopy restoration and other non-point source control efforts would produce far greater ecological benefits at far lower costs. This analysis explores EPA's historical approach to nutrient regulation in streams and explains why prior efforts were unsuccessful. It also provides a critique of the agency's latest proposal with the hope that a more effective and comprehensive approach to stream restoration may be established.

Background

Requirements for Water Quality Standards

By statute, criteria must be based on the latest available science and set at the level necessary to protect aquatic life and human health uses.[2] To achieve this requirement it is essential that criteria possess two attributes: (1) the criteria must be based on data that confirm the pollutant is causing use impairment at given concentrations, and (2) the level at which the numeric criteria is set is both sufficient and necessary to protect stream uses. Thus, criteria are, in general, set at the threshold level where the pollutant exposure is demonstrated not to pose a significant threat to aquatic life.[3]

Since 1985, EPA has had a well defined procedure for developing water quality criteria when it published the *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses* (1985) (hereafter *Guidelines*).[4] The *Guidelines* establish a number of very specific scientific requirements for the establishment of numeric criteria germane to EPA's new weight-of-evidence approach:

- All decisions should be based on a thorough knowledge of aquatic toxicology and criteria decisions must be altered if there is a substantial probability of over or under protection of aquatic organisms and their uses.[5]
- It is not enough that the criterion is the best estimate given the available data. Criteria should be derived ''only if adequate appropriate data are available to provide reasonable confidence that it is a good estimate.''[6]
- Criteria must be based upon studies showing a clear dose/response relationship to determine effect concentration. Data from confounded studies (i.e., results that are influenced by factors other than the pollutant of concern) should not be used.[7]
- Where the effects of a pollutant are shown to be related to other water quality characteristics, the standard should be derived with consideration for that water quality characteristic.[8]
- Based on ''all available laboratory and field information'', it must be determined that proposed criteria are ''consistent with sound scientific evidence.'' If not, another criterion should be derived.[9]

A numeric water quality standard represents the quality of water that supports a particular use. It is well recognized that use impairment is not caused directly by nitrogen or phosphorus.[10] Rather, use impairment results from nuisance levels of algae and other alterations in aquatic vegetation that may develop in response to nutrient enrichment.[11] EPA reiterated this primary concern in its updated national nutrient strategy in May 2007. However, excess nutrients do not always stimulate excessive plant growth. Many physical factors and several biological factors either allow or prevent nutrients from stimulating excessive plant growth.[12] Thus, it is essential to understand the conflu-

[2] Clean Water Act Section 304(a)

[3] Clean Water Act Section 304(a); 40 CFR 131.2 131.3 (b), (c)

[4] This document is available at http://www.epa.gov/waterscience/criteria/library/85guidelines.pdf.

[5] *Guidelines* at 5, 18.

[6] *Id.* at 5.

[7] *Id.* at 15, 16, 21.

[8] *Id.* at 29, 40.

[9] The concluding recommendation of the *Guidelines* at 57.

[10] *Nutrient Criteria Technical Guidance Manual: Rivers and Streams*, EPA-822-B-00-002 (July 2000), hereafter, the *Rivers and Streams Document*, available at http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/rivers-streams-full.pdf.

[11] *Id.* at 4.

[12] Plant growth can become excessive, even under nutrient poor conditions, when canopy is removed and streams are exposed to sunlight. Well shaded streams may not exhibit excessive plant growth, even when nutrient concentrations are high; zebra mussels populations (aggressive filter feeders) have sig-

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

ence of factors that control excessive plant growth in different ecosystems to develop effective restoration measures.

Based on the *Guidelines* requirements and the manner in which nutrients affect designated uses, one would expect that any numeric nutrient water quality criteria must be based on the stimulation of plant growth. The level of plant growth that causes designated use impairments, whether to aquatic life, drinking water, or recreational uses, would need to be defined. Then, the nutrient concentrations necessary to support such growth would be identified and confounding factors that influence that relationship, such as shading, biological influences, scour or other relevant physical factors would be taken into account. Alternatively, the level of plant growth associated with the use impairment could become the water quality standard and confounding factors could be addressed through the TMDL process, thus bypassing the difficult task of relating nutrient levels and plant growth. EPA's guidance, discussed below, recommends such approaches.

EPA Guidance

EPA has developed extensive guidance on developing numeric nutrient criteria and related TMDLs. For example, see:

- *Protocol for Developing Nutrient TMDLs: First Edition*, EPA 841-B-99-007 (November 1999), available at http://www.epa.gov/owow/tmdl/nutrient/pdf/nutrient.pdf.
- *Nutrient Criteria Technical Guidance Manual: Rivers and Streams*, EPA-822-B-00-002 (July 2000), available at http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/rivers-streams-full.pdf.
- *Ambient Water Quality Criteria Recommendations: Information Supporting the Development of State and Tribal Nutrient Criteria for Rivers and Streams in Nutrient Ecoregions I - XIV* (December 2000), available at http://www.epa.gov/waterscience/criteria/library/.

The basic scientific premise underlying all published EPA nutrient criteria development documents is that nutrient control is intended to reduce excessive plant growth that results in use impairment. Consistent with the *Guidelines*' requirements for a clear demonstration of causation, the various EPA nutrient criteria documents all specify that dose/response demonstrations are required to set scientifically defensible nutrient standards. Nutrient levels must be documented to ''cause'' specific changes in plant growth (typically measured as chlorophyll 'a') and other physical variables directly affected by excessive plant growth (secchi depth, DO, transparency, etc.).[13]

> Nutrient criteria development should relate nutrient concentrations in streams, algal biomass and changes in ecological condition (e.g., nuisance algae accrual rate and deoxygenation). . . . Initial criteria should be verified and calibrated by comparing criteria in the system of study to nutrients, Chl-a and turbidity values in water bodies of known condition to ensure that the system of interest operates as expected.[14]
>
> Predictive relationships between nutrients and periphyton (or phytoplankton) biomass are *required* to identify the critical or threshold concentrations that produce nuisance algal biomass.[15]

The various EPA nutrient criteria development documents also acknowledge that nutrients may cause ecosystem impacts to upper level organisms (invertebrates, fishes), but never directly:

> However, fish and macroinvertebrates do not directly respond to nutrients, and therefore may not be as sensitive to changes in nutrient concentrations as algal assemblages. It is recommended that relations between biotic integrity of algal assemblages and nutrients be defined and then related to biotic integrity of macroinvertebrate and fish assemblages in a stepwise, mechanistic fashion.[16]

EPA's published guidance indicates that invertebrate populations may be impacted only when plant growth rises to a level where extensive/excessive plant growth causes those ecosystem changes. These changes are not documented to occur directly due to nutrients as this parameter is not a toxicant and does not have a direct impact on sensitive organisms.[17] This fact also was well documented by EPA's field studies under the whole effluent toxicity program. Thus, proper criterion development requires confirmation that (1) nutrients are causing excessive plant growth in the system of interest and (2) the growth is above some acceptable level that will cause environmental harm.

While phosphorus and nitrogen have been targeted as the agents responsible for excessive plant growth, nutrients are not the sole determinants of plant growth, as noted in a recent EPA document to support several stringent nutrient TMDLs:

> Study results summarized as part of this literature review support the assertion that while a relationship may exist between periphyton growth and nutrients, the dynamics change as a function of multiple factors. These factors include antecedent conditions, water temperature, pH, light availability, flow regime, and grazing, among others. Nutrient levels may be secondary to other determinants of biomass and growth such as light, disturbance, and grazing.[18]

Over the last decade, EPA and the states have made limited progress toward developing numeric nutrient

nificantly reduced the level of phytoplankton in Lake Erie, but the loss of oyster populations in the Chesapeake Bay is largely responsible for the eutrophication problems experienced there.

[13] *See, Rivers and Streams Document*.

[14] *Id*. at 13.

[15] *Id*. at 76 (emphasis supplied).

[16] *Id*. at 85.

[17] *See*, *Development and Adoption of Nutrient Criteria into Water Quality Standards* (2001) at 14, response 4, available at http://www.epa.gov/waterscience/criteria/nutrient/files/nutrientswqsmemo.pdf.

[18] *Literature Review to Support Selection of Nutrient TMDL Endpoints for Northern Piedmont Ecoregion Streams in Southeastern Pennsylvania* (January 2008) Prepared by Tetra Tech Inc. for EPA Region III., from *Response Document for Nutrient and Sediment TMDLs in Pennsylvania for Southampton Creek, Indian Creek, Chester Creek, Paxton Creek and Sawmill Run* (June 30, 2008) available at http://www.epa.gov/reg3wapd/tmdl/pa_tmdl/NutrientEndPoint/Response_Document_Part_A.pdf.

water quality standards for streams. In 2001, EPA published recommended water quality criteria, based on an eco-regional evaluation of ''reference conditions'' wherein EPA presumed that meeting reference stream nutrient concentrations would prevent adverse effects for all designated uses. EPA developed these criteria with the intention that they serve as a ''*starting point*,'' from which the states would develop more refined ''effects based'' nutrient criteria using other scientifically defensible approaches. In a 2001 memorandum to the states, EPA urged the states to adopt their own nutrient water quality standards.[19] Many states began their own nutrient impact assessments, in the hope of identifying threshold levels for nutrient impairment; however, due to the scientific complexity, few identified appropriate stream objectives. On May 25, 2007, , EPA sent a memorandum to the states urging that they take ''bold steps'' in adopting numeric nutrient water quality criteria.[20] However, the simplistic approaches specified in federal guidance faced a harsh reality—nutrient control was not always the answer. Several leading scientists had informed EPA of this fact, but to no avail.

State Concerns with Federal Approach

Following years of analyses using EPA's recommended approaches, on July 18, 2007, the Association of State and Interstate Water Pollution Control Administrators (ASIWPCA) submitted a letter to EPA's Office of Water, concerning the development of nutrient water quality standards.[21] This letter was precipitated by EPA's May 25, 2007, memorandum which urged accelerated promulgation of numeric nutrient water quality criteria by the states. In part, ASIWPCA's comments rebutted EPA's suggestion that the states were not being diligent with regard to nutrient criteria development. More importantly, this letter highlighted the problems that states encountered in developing scientifically defensible numeric nutrient standards:

> During their considerable development processes, many States are failing to find a strong linkage between the EPA recommended cause variables (N and P) and response variables of chlorophyll-a and transparency, but are finding wide variations in parameters that seem unrelated to professional assessments of ''trophic health'' status. In many cases, a relationship cannot be demonstrated between causal variables N and P, and factors such as turbidity, light limitation, canopy cover, substrate, aquatic community structure, bioavailability, reservoir sequestration, micronutrient limitations and other ''response'' variables. These problems can only lead to mis-cues in impairment identification and mis-direction of scarce management and implementation resources.

Consequently, ASIWPCA asked EPA to continue to refine and enhance the scientific basis for numeric nutrient criteria, including reevaluating the potential difficulties with statistically-derived generic criteria that may be over or under protective. The ASIWPCA letter even suggested that technology-based requirements for nutrients may be a more appropriate approach. In November 2007, the NRDC and others filed a petition for rulemaking with EPA[22] to incorporate nutrient removal (state-of-the-art nitrogen and phosphorus removal) into the secondary treatment standards and bypass the need to demonstrate such limitations are necessary to protect the environment as the act currently requires.[23]

EPA Unveils 'Weight-of-Evidence' Approach

In December 2007 EPA unveiled a radical new approach to establishing nutrient endpoints in TMDLs. This approach applied a ''weight of evidence'' analysis to derive instream nutrient endpoints that are not associated with any site-specific designated use impairments. The weight-of-evidence approach was used to set endpoints for total phosphorus (TP) in five watersheds across Pennsylvania.[24] These endpoints, ranging from 25 – 40 μg TP/L as growing season averages, were by far the most restrictive endpoints developed to date in these watersheds and will require nutrient reductions up to 90 percent from stormwater outfalls and in excess of 98 percent from POTWs. The cost for compliance is staggering and, in the case of MS4 (storm runoff) discharges, is likely unachievable. The weight-of-evidence analysis also was used to develop TN reduction requirements; although the TMDL analyses did not indicate TN was affecting plant growth or invertebrate populations.

EPA has confirmed that it intends to use this approach nationwide. A June 11, 2008, memorandum stated that the ''multiple lines of evidence'' (or weight-of-evidence) approach was scientifically defensible and consistent with EPA guidance for deriving nutrient criteria. [25] This approach is featured in EPA's N-STEPS Web site and, as such, is being made available for nationwide application and has already been applied by EPA to assess potential stream standards for Illinois[26]. In August 2008, the group of Pennsylvania municipalities affected by the new criteria development methodology filed a request for SAB Peer Review.[27] The request was supported by more than 20 public and private organizations. In December 2008, EPA agreed to the peer review, but made it clear the agency intended to use the

[19] Memorandum from Geoffrey Grubbs, *Development and Adoption of Nutrient Criteria into Water Quality Standards* (Nov. 14, 2001), available at http://www.epa.gov/waterscience/criteria/nutrient/files/nutrientswqsmemo.pdf.

[20] Memorandum from Benjamin H. Grumbles, *Nutrient Pollution and Numeric Water Quality Standards* (May 25, 2007), available at http://www.epa.gov/waterscience/criteria/nutrient/files/policy20070525.pdf.

[21] Letter from the Association of State and Interstate Water Pollution Control Administrators (ASIWPCA) to Ben Grumbles (July 18, 2007) available at http://www.asiwpca.org/home/docs/Ltr2EPANutrients.pdf.

[22] *Petition for Rulemaking under the Clean Water Act—Secondary Treatment Standards for Nutrient Removal* (Nov. 27, 2007), available at http://elpc.org/wp-content/uploads/2008/08/nutrientpetitionfinal.pdf.

[23] *See*, Clean Water Act § 301(b)(1)(C).

[24] *See Nutrient TMDLs for Goose Creek, Indian Creek Watershed, Paxton Creek Watershed, Sawmill Run, and Southampton Creek* (June 30, 2008) available at http://www.epa.gov/reg3wapd/tmdl/.

[25] Memorandum from William Swietlik, Chief, Ecological and Health Processes Branch, EPA Office of Water to Robert Koroncai, Associate Director, Water Protection Agency, EPA Region III, *Development of Nutrient Endpoints for TMDLs in Pennsylvania* (June 11, 2008).

[26] Tetra Tech Inc., *Data Analysis Report for Analysis of Illinois Stream and River Nutrient and Biological Data for the Nutrient Scientific Technical Exchange Partnership Support (N-STEPS)* (Oct. 2, 2008).

[27] Letter from John Hall to Stephen Johnson, USEPA Administrator, *Request for Peer Review of New EPA Region III Approach to Developing Instream Standards for Nutrients* (Aug. 21, 2008).

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

new nutrient criteria derivation methodology nationwide.

The cost and environmental implications of the new methodology are staggering. Several hundred billion dollars in nutrient controls will be dictated by such analyses. Stormwater limits will be unattainable and therefore may require that no growth be allowed in many watersheds. Costs notwithstanding, because the new approach does not require a direct connection to environmental need or a showing that environmental protection results from the method, such resource expenditures may not produce discernable environmental benefits. Consequently, decades may pass before the necessary environmental improvements are initiated. To understand why the peer review request was filed, the reasons behind EPA's development of this methodology must be reviewed. The development of the weight-of-evidence approach followed several failed alternative approaches to nutrient standard development. These approaches, and the problems encountered in their application, are discussed below.

Nutrient TMDL Evolution: Problems, Failures

It is fairly well understood that nutrients, combined with a multitude of other factors, may stimulate excessive plant growth. As the plant biomass increases, it results in pronounced diurnal variation in dissolved oxygen levels in response to the photosynthesis – respiration cycle, with the primary concern being DO depletion prior to dawn, when the effects of plant respiration reach a peak. Low DO levels may significantly impair a host of sensitive vertebrate and invertebrate species. In medium to small streams, these variations are caused by attached plant growth (periphyton and macrophytes), *not* floating algae (phytoplankton). Although the underlying science is understood, implementation of effective nutrient TMDLs in river and stream systems was still uncertain because the dynamics of periphyton and macrophyte growth are very complicated and good predictive models are not available, as acknowledged in EPA's guidance documents:

> [D]eveloping predictive relationships between nutrient load and periphyton biomass can present considerable challenges because well-known and validated water quality models are not available.[28]

> Macrophytes depend primarily on sediments for nutrient uptake, and are relatively unaffected by nutrient water column concentrations. However, attempts to relate macrophyte growth or biomass with sediment nutrient content have been largely unsuccessful.[29]

Nonetheless, EPA's published guidance, citing publications from Dodds and Welsh, assumed periphyton levels would be affected directly by reducing nitrogen and phosphorus levels. Apparently, EPA presumed periphyton respond like phytoplankton to nutrient controls. Based on this assumption, EPA began preparing TMDLs in 2005 that regulated phosphorus in an effort to control ''nuisance'' periphyton growth in several Pennsylvania streams. The approach used in these TMDL efforts was eventually shown to be misplaced and all of these TMDLs were either withdrawn or not finalized.[30] Eventually, it came to be understood that periphyton do not respond like phytoplankton. The stream data confirmed high periphyton growth even at very low nutrient levels if other environmental factors were favorable for plant growth (e.g., light, substrate). While this was happening in Pennsylvania, other states attempted similar approaches or pushed ahead with EPA's distribution/regression-based approach (e.g., Montana). EPA began receiving numerous complaints that the methods recommended in the published criteria development documents just did not work. Consequently, in 2008, EPA unveiled a radically new approach that divorced nutrient water quality standards from plant growth. This approach apparently was developed to jump over the difficulty states were encountering in attempting to derive plant growth-nutrient relationships. This weight of evidence approach currently is being promoted as the ''state of the art'' approach to nutrient criteria development.[31] The evolution of these approaches and the context in which they were applied is detailed below.

Nuisance Algal Growth

■ Pennsylvania Experience (Generic Regressions Don't Work)

From 2005-2008, EPA Region III and the Pennsylvania DEP began development of several nutrient TMDLs, targeting ''nuisance algae'' on effluent-dominated streams in southeastern Pennsylvania in 2005 to control periphyton growth and *Cladophora* (a filamentous green algae).[32] Initially, these TMDLs used an interpretation of the state's narrative standard to set a threshold of 100 mg Chl-a/m^2 (growing season average) as the ''nuisance algae'' threshold, above which uses were impaired.[33] Literature data from a ''world database'' was used in a regression analysis[34] to equate the ''nuisance algae'' level with a maximum instream phosphorus concentration. This resulted in an endpoint estimate of ~0.20 mg/L for total phosphorus, although the regression factor indicated that over 80 percent of the variability in periphyton biomass was attributed to factors *un-*

[28] *Nutrient TMDL Protocol* at 4-6.

[29] *Rivers and Steams Document* at 73.

[30] The Skippack Creek TMDL was prepared by EPA and finalized in 2005 and withdrawn in 2008. The Pennsylvania Department of Environmental Protection prepared a similar draft TMDL for the Neshaminy Creek in 2006 which was never finalized. EPA also prepared a draft TMDL for Wissahickon Creek in 2006 that was withdrawn.

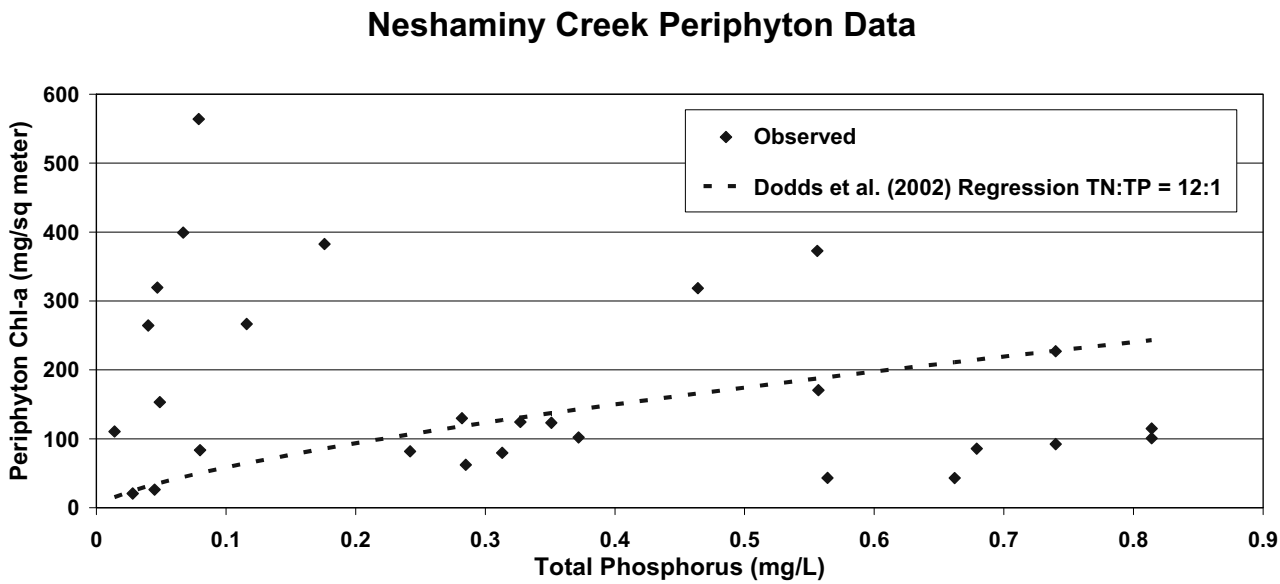
[31] The more accurate description is that this is a step backwards to the Dark Ages, prior to the time when empirical evidence was needed to confirm scientific conclusions.

[32] A 1997 EPA analysis of the Clark Fork River in Montana determined that *Cladophora* could not be controlled. Apparently, *Cladophora* were found abundant in nutrient poor waters but largely absent in nutrient rich waters. Dodds, W.K., V.H. Smith, and B. Zander, *Developing Nutrient Targets to Control Benthic Chlorophyll Levels in Streams: A Case Study of the Clark Fork River*. Water Research 31(7):1738 – 1750 (1997).

[33] Carrick, H.J. and C.M. Godwin, *TMDL Endpoint Estimates for an Urban-Suburban Stream Based Upon In-stream Periphyton Assemblages (Neshaminy Creek, Pennsylvania): Final Report* (Submitted to the Pennsylvania Department of Environmental Protection, Dec. 15, 2005).

[34] Dodds, W.K., V.H. Smith, and K. Lohman, *Nitrogen and Phosphorus Relationships to Benthic Algal Biomass in Temperate Streams*, Can. J. Fish. Aquat. Sci. 59:865-874 (2002).

Figure 1

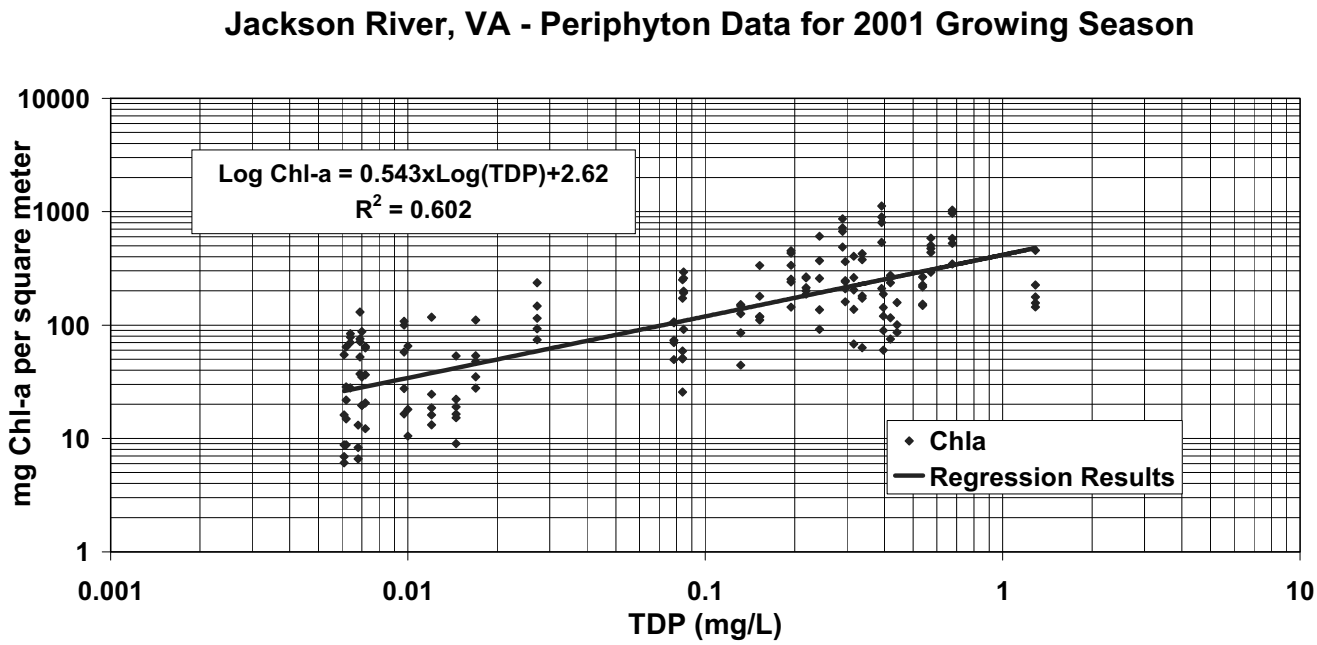


related to nutrient concentration (i.e., in all likelihood, this approach would not reduce plant growth or restore uses). Using this approach, EPA established a TMDL for the Skippack Creek (2005). The same approach was used to develop draft TMDLs for Neshaminy Creek (DEP, 2006) and Wissahickon Creek (EPA, 2006). The approach was premised on the assumption that *Cladophora* growth would be limited under these conditions although *none* of the projections for reduced plant growth were based on site-specific information.[35]

[35] TMDL rules require the TMDL to be based on site-specific data. 40 CFR 130.7 (c) (1).

The primary problem with this approach was that (1) ''nuisance'' periphyton thresholds for use impairment have never been defined in Pennsylvania's water quality standards and (2) the predicted benefits of nutrient control were at odds with the available site-specific data. Although the Pennsylvania DEP claimed that the watersheds were impaired by nutrients for *aquatic life*, the algal threshold (100 mg Chl-a/m^2) was traced in the literature as an *aesthetics issue* (an impediment to fisherman foot traffic if the algae exceeded 150 mg Chl-a/m^2 and exceeded 20 percent *filamentous* forms). Thus, the nuisance algal threshold was not related to aquatic life impairment nor was it apparent how this level of plant growth compared with natural conditions

Figure 2



 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

in Pennsylvania that exceed these levels. Data on *Cladophora* showed that even in the cleanest of waters (10 – 20 μg TP/L) this plant could thrive if other conditions were favorable. Moreover, the ''world database'' regression relationship used to establish the TP endpoint did not match the site-specific data for any of the watersheds, as illustrated in **Figure 1**. Consequently, the endpoint level selected (0.2 mg TP/L) could not possibly limit plant growth to targeted levels. Some of the highest periphyton biomass levels were observed in conjunction with much lower TP concentrations occurring *above* all wastewater inputs. For unexplained reasons, although EPA/DEP had embarked on an extensive site-specific data collection campaign, the regulatory agencies chose to ignore that information in relying on the ''Dodds Regression.''

These various TMDLs were eventually withdrawn, not because the model obviously did not fit the site specific data, but because the author of the original regression equation (Dodds et al., 2002) subsequently published an addendum (Dodds, 2006), citing errors in the original data base and revising the regression.

■ Jackson River, Virginia (Site-Specific Regressions Also Don't Work)

In a similar effort, also using EPA's recommended regression approach, the Virginia DEQ developed a 2006 nutrient control strategy for the Jackson River. The goal was to reduce periphyton growth to address dissolved oxygen and biotic index impairments. First, a literature survey was completed, indicating a periphyton—TP relationship could be developed. In this case, unlike the Pennsylvania TMDLs, a site-specific regression equation was developed based on detailed sampling to quantify periphyton growth in relation to TP exposure in the Jackson River. The regression was developed from an extensive database, collected by VDEQ in 2001, which indicated a high correlation between total dissolved phosphorus and periphyton biomass (R^2 = 0.6). In this respect, the analysis represented a significant improvement over the earlier ''world database'' regression analysis and suggested that site-specific data were necessary to develop useful relationships (**Figure 2**). The data analysis indicated that controlling TP would cause significant reductions in periphyton growth. The surprising aspect of this analysis was the very high correlation coefficient, which indicated a much greater periphyton growth dependency on TP levels than prior published values. If accurate, this would have been a reasonable basis for TP reduction in these waters.

The state's narrative water quality standard was interpreted to set a periphyton biomass target of 80 mg Chl-a/m^2 to ensure meeting the growing season average 100 mg Chl-a/m^2 impairment threshold. The regression yielded a growing season total dissolved phosphorus (TDP) endpoint of 0.047 mg/L (equivalent to 0.063 mg TP/L). The TMDL was implemented and periphyton growth was subsequently evaluated in 2006. The site-specific data for 2006 demonstrated that, although instream TP now averaged about 0.02 mg/L, well below the regression-based TP target, there was *no* material change in the periphyton biomass between 2001 and 2006. A subsequent review of the original regression indicated that certain data, available in 2001, were overlooked[36] and that the manner in which the data were evaluated created an impression of greater certainty than actually existed. A closer review of the data indicated that the range of periphyton growth, in response to TP concentrations greater than 20 μg/L, is not meaningfully different, regardless of TP concentration. Moreover, most researchers evaluate periphyton biomass and nutrient levels on a growing season average basis, as opposed to the individual observations presented in **Figure 2**. This helps to dampen out data scatter which is not related to changing nutrient levels. The growing season averages for the 2001 and 2006 data are presented in **Figure 3** to illustrate that the TP reduction did not have any effect on periphyton growth.

[36] Data left out of the original regression analysis showed chlorophyll-a levels well above 200 mg/m2 in associated with TDP levels below 0.01 mg/L.

Figure 3

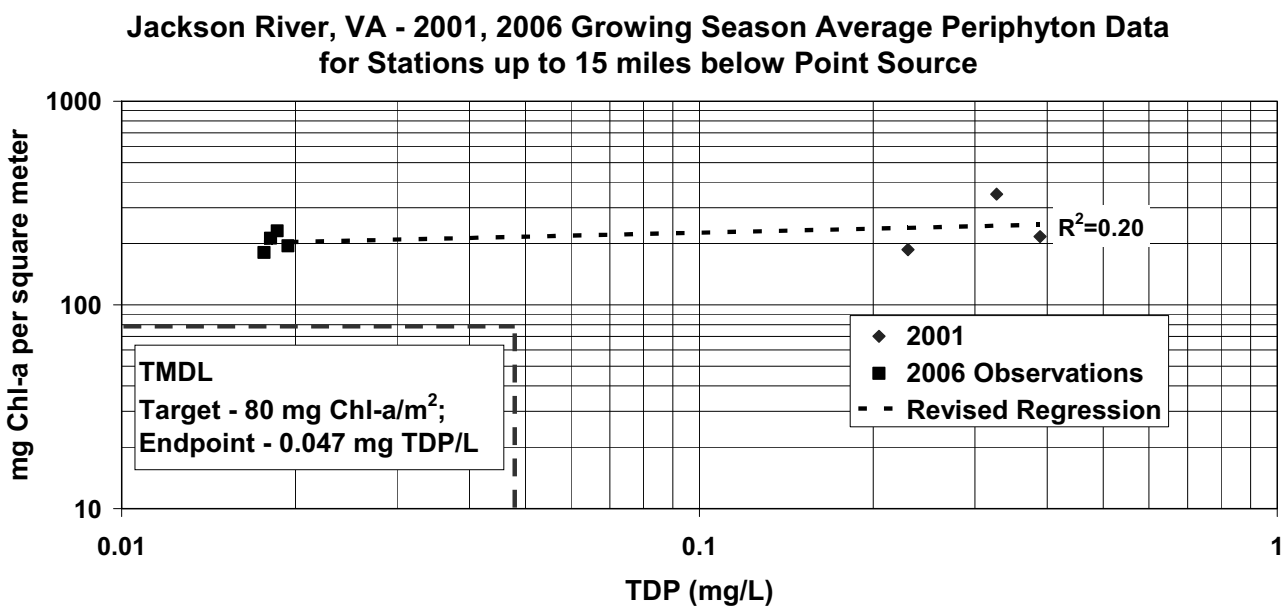


Figure 3 shows that the significant reductions in TDP, over an order of magnitude and well below the regression-based endpoint estimate of 0.047 mg TDP/L, did not result in any significant change in periphyton biomass. Although the regression model predicted a periphyton chlorophyll-a concentration of 80 mg/m^2 at 0.047 mg TPD/L, the results indicate that periphyton levels exceeded 200 mg Chl-a/m^2 with TDP levels less than 0.020 mg/L (growing season average). Clearly, periphyton were not responding to nutrient reduction as anticipated and the intended benefits of nutrient regulation – increased minimum DO, biotic index improvement – did not occur.

■ East Canyon Creek, Utah

Another nutrient TMDL, originally prepared by an EPA Contractor in 2000 for East Canyon Creek in Utah, was developed to address dissolved oxygen impairments attributed to macrophyte (e.g., rooted, vascular plants) growth in the stream.[37] The Utah DEQ did not have stream nutrient standards to apply in the TMDL and elected to use an interim TP concentration of 0.05 mg/L until the assimilative capacity of the creek was better understood. To comply with the limits specified in the TMDL, Park City (the only point source on the stream) implemented a $12 million upgrade to provide biological and chemical nutrient removal for the wastewater treatment facility. After these upgrades, the facility effluent averaged 0.04 mg/L TP in 2004/2005 and effluent quality was generally equal to or better than the ambient stream water quality. In spite of these improvements, the stream still experienced dissolved oxygen impairments and extensive macrophyte growth. Subsequent research, completed by the University of Utah, confirmed that macrophytes rely primarily on the sediment for nutrients. Thus, changes in overlying water column levels of phosphorus are largely irrelevant.

This result should have been expected given that the dissolved oxygen impairment was attributed to macrophytes. As noted previously, macrophytes are able to obtain nutrients from the sediment[38] and water column nutrient concentration is unlikely to affect their growth. The dominance of macrophytes in this system can be attributed to sedimentation that has created a favorable habitat for these plants. Without this habitat, the stream would likely support robust periphyton growth similar to that seen in the Jackson River. Finally, like the periphyton situation in Pennsylvania, macrophyte growth *above* the Park City facility was often as robust as that occurring below the facility. This should have alerted the consultant preparing the TMDL that regulating the point source would likely produce no discernible benefit. However, EPA's single-minded focus on nutrient reduction as the solution to all plant growth problems seems to blind Agency consultants to otherwise contradictory information.

■ Wenatchee River, Washington

In 2008, the Washington Department of Ecology released a draft nutrient TMDL for the Wenatchee River to address dissolved oxygen and pH variations associated with periphyton growth. The Wenatchee is a fairly large river with a typical summer average flow greater than 1,000 cubic feet per second (cfs), and a late growing season (July—October) 7Q10 flow of 344 cfs. Three relatively small wastewater facilities discharge in the basin, with design flows ranging from 0.1 – 1.9 million gallons per day (MGD). A review of the data used to support the TMDL findings indicates that impressive amounts of periphyton biomass develop under exceeding low levels of nutrients. The Washington Department of Ecology provided information indicating that upwards of 35 g AFDM/m^2 (ash-free dry mass; equivalent to about 150 – 300 mg Chl-a/m^2) periphyton biomass developed in response to ortho-phosphorus concentrations averaging about 0.004 mg/L.[39] Moreover, periphyton growth did not appear to respond to increases in instream ortho-phosphorus concentration beyond that observed due to natural background conditions. These observations further support the conclusion that periphyton growth reaches high levels in response to extremely low levels of phosphorus, and point source control may not be effective in reducing periphyton growth below the outfalls.

This is not unexpected since periphyton consist of a community of aquatic plants that compete for resources and adapt to changing environmental circumstances. Several years after publishing his world data base regression, Dodds[40] noted ''attached algae might be able to attain impressive biomass in nutrient poor water because periphyton can use the small amounts of nutrients that continuously flow by.'' This ability is quite obvious in the Wenatchee River. Thus, it is apparent that regulatory agency attempts to restrict periphyton growth have misunderstood how these organisms respond to nutrient levels and what factors control their growth. Nonetheless, DOE proposed TP reductions of 99 percent from the point sources.

■ Summary on Nuisance Algal Growth Approaches

As discussed above, regression analyses to related periphyton growth (as Chlorophyll-a) to TP have not been successful in their ability to predict how to regulate such growth. If excessive plant growth is not reduced, the secondary impacts of such growth – low minimum DO, altered invertebrate populations – will not be improved. The available data suggest that periphyton can achieve high densities in response to extremely low TP concentrations, as illustrated in the Neshaminy Creek, the Jackson River, and the Wenatchee River. Based on these results, TP levels would need to be well below 10 μg/L to limit such growth if growing conditions are favorable. Alternatively, the data also indicate that other conditions, unrelated to nutrients, may be far more effective in limiting periphyton growth, even when TP concentrations are well above 100 μg/L. (See Figures 4 and 5)

Measurements made by the Pennsylvania DEP on Wissahickon Creek illustrate the overriding influence of sunlight on periphyton growth (**Figure 4**). When the tree canopy along the stream bank is intact, with

[37] Utah Department of Environmental Quality, Division of Water Quality, *Total Maximum Daily Load for East Canyon Creek*, (April 1, 2000), available at http://www.waterquality.utah.gov/TMDL/East_Canyon_Creek_TMDL.pdf.

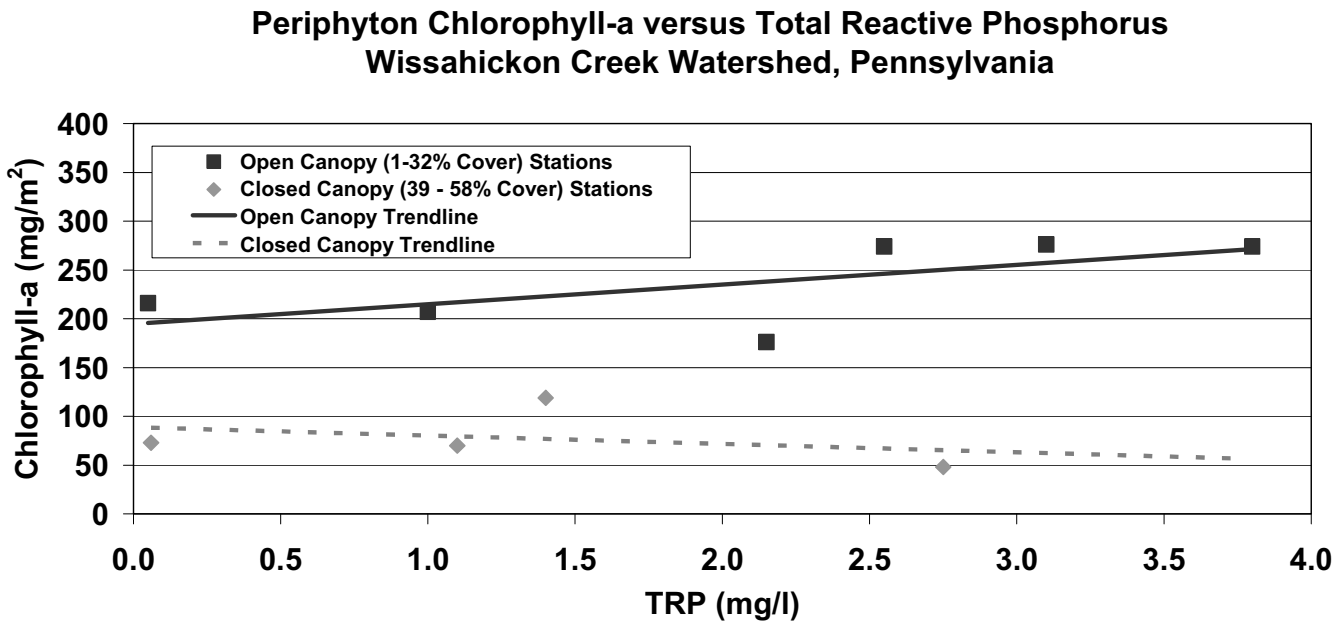
[38] *Rivers and Streams Document* at 73.

[39] *See*, http://www.ecy.wa.gov/programs/wq/tmdl/WenatcheeMulti/DOpH.html.

[40] Walter K. Dodds, *Eutrophication and Trophic State in Rivers and Streams*, Limnol. Oceanogr. 51(1, part 2) p. 671-680 (2006), available at .http://www.aslo.org/lo/toc/vol_51/issue_1_part_2/0671.pdf.

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

Figure 4

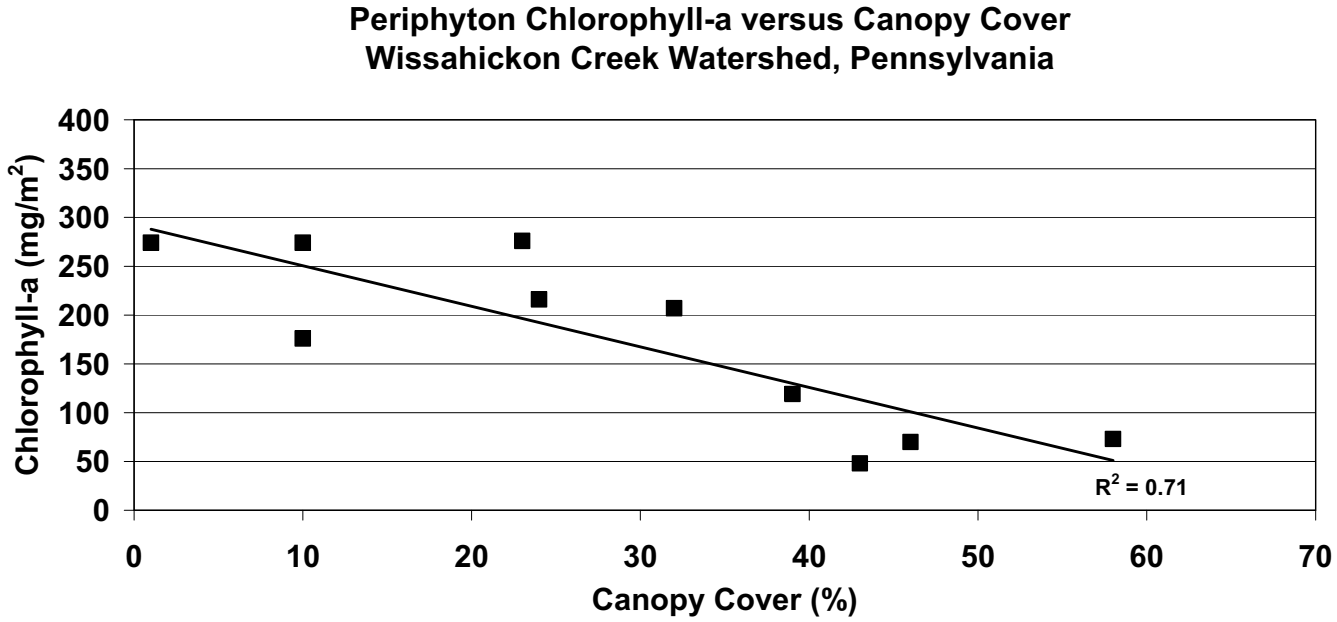


canopy coverage greater than 39 percent, periphyton growth is not excessive, even when phosphorus concentrations exceed 1,000 μg/L. However, when the canopy cover is 32 percent or less, periphyton growth generally exceeds 200 mg Chl-a/m^2. The same data, when plotted as a function of canopy cover, show a strong relationship between the amount of canopy cover and periphyton growth (**Figure 5**). This relationship is much stronger than any presented for TP, suggesting that canopy restoration would be much more effective as a mitigation approach to reduce plant growth in streams.

Distribution-based Endpoint (Assumed Impairment Approach)

Following one of EPA's recommended approaches, the Montana Department of Environmental Quality (DEQ) recently prepared a document outlining the Department's approach to developing statewide nutrient

Figure 5



criteria.[41] This document presents a detailed evaluation of ecoregions within the State, identifies reference streams within those ecoregions, and presents statistical distributions for the data similar to the distribution-based approach presented by EPA in the 2001 *Rivers and Streams Document*. DEQ then considered stressor-response study data to determine a target criterion and associated that target with its corresponding percentile from the distribution data. In the case of the mountainous ecoregions, DEQ set a daily maximum instream benthic algae (periphyton) concentration of 150 mg Chl-a/m^2 and 36 g AFDM/m^2 as a use impairment threshold, based on a public perception survey on *recreational aesthetics*. DEQ then derived an instream phosphorus concentration associated with the public perception impairment threshold from literature studies (0.010 mg soluble reactive phosphorus (SRP)/L). These literature studies used regression-type analyses to associate phosphorus concentrations with the target periphyton concentrations. The regression-based phosphorus concentration was equivalent to the 90th percentile TP concentration for reference streams. Consequently, DEQ recommended TP, TN, and nitrate-nitrogen criteria equal to the 90th percentile of the reference stream distributions (**Table 1**).

Table 1
Nuisance Algae-based Numeric Nutrient Criteria (mg/L)

Level III Ecoregion	TP	TN	NO_3
Northern Rockies	0.012	0.233	0.081
Canadian Rockies	0.006	0.209	0.020
Middle Rockies	0.048	0.320	0.100
Idaho Batholith	0.011	0.130	0.049

In the case of the eastern Montana prairie streams, the department attempted to generally relate compliance with existing water quality criteria (dissolved oxygen) to nutrient level as the basis for its ''stressor-response'' evaluation. The analysis, however, did not look at dissolved oxygen. Rather, it looked at diatom communities and inferred compliance with the DO standards based on a diatom oxygen tolerance index (OTI). Finally, DEQ also evaluated diatom OTI versus TN concentration to define the threshold TN level, above which, DO impairments might be inferred to exist. This regression-based TN threshold was near the 75th percentile of the reference stream distribution. Consequently, DEQ recommended nutrient criteria based on the 75th percentile concentrations (**Table 2**).

Table 2
DO-inferred Numeric Nutrient Criteria (mg/L)

Level III Ecoregion	TP	TN	NO_3
Northwestern Glaciated Plains	0.123	1.311	0.020
Northwestern Great Plains	0.124	1.358	0.076

Notwithstanding the arbitrary assumption that, once a reference stream percentile is established for one parameter, it can be applied to all other nutrient parameters of concern as necessary to ensure use protection, this approach is only reasonable if the linkage between use impairment and nutrient level is clearly established. However, DEQ's approach depends entirely on the stressor-response relationships and assumes various impairment relationships. Presume, for example, that the public perception periphyton concentration (150 mg Chl-a/m^2) is an appropriate use impairment threshold. The literature relationships cited as the basis for defining phosphorus level, above which the periphyton threshold is exceeded, is based on a regression. These regressions have a high degree of error associated with them and have been shown in other situations to be useless for deriving TMDL endpoints[42] The diatom inferred oxygen concentration is even more esoteric and prone to error. Moreover, regulating one nutrient (e.g., TN, nitrate) because TP requires control is not a rational scientific approach and greatly increases the cost and energy usage for regulation plant growth.

The Montana approach shows that the consequence of setting water quality standards based on a series of tenuous assumptions (i.e., setting a TN standard based on the TP percentile) can be enormous, particularly for smaller communities. For example, Philipsburg, Mont. is located in the Middle Rockies Level III Ecoregion, has a population of 930 residents (about 450 households), and wastewater from the town's lagoon system discharges to a small stream, Flint Creek. DEQ has estimated that, to comply with the nutrient standards, Philipsburg will need to replace the simple-to-operate lagoon system with a new, mechanical BNR facility. The cost of this replacement is estimated at $6.1 million and per household sewer rates will rise from approximately $15 per month to more than $70 per month, nearly a 400 percent increase. This increase represents more than 2 percent of the median household income in Philipsburg. Consequently, it is likely to result in economic distress for the town residents. Moreover, the new treatment facility will still *not* be able to achieve the TN and nitrate-nitrite limits that will result from imposing water quality standards of 0.32 mg/L and 0.10 mg/L, respectively. Only zero discharge facilities could be expected to comply with such effluent quality requirements.

Perhaps more disturbing is the likelihood that the designated use impairment will not be corrected by these exorbitantly expensive measures. The ''Ecoregion'' TP criterion for the Middle Rockies ecoregion is 0.048 mg TP/L. The stressor-response evaluation used to relate periphyton chlorophyll-a to phosphorus concentration was attributed to several authors, including Welch et al.[43] Welch reported ''there is little hope that biomass at any stream point could be controlled by controlling ambient SRP'' for his study of the Spokane River. This study reported that ''nuisance'' biomass levels of 150 mg Chl-a/m^2 would be exceeded in the Spokane River over distance of 13 – 20 kilometers at an SRP of 0.010 mg/L. But, the proposed criterion is potentially five times higher (0.048 mg TP/L) and DEQ intends to apply the 150 mg Chl-a/m^2 threshold as a daily maximum value. Based on the more current work of Dodds and Welsh, there is little likelihood that periphy-

[41] Suplee, M, V. Watson, A. Varghese, and J. Cleland, *Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers* (November 2008), available at http://www.deq.state.mt.us/wqinfo/Standards/WhitePaper_FNL3_Nov12-08.pdf.

[42] *See, supra* discussions regarding Nuisance Algal Growth.

[43] Welch, E.B., Horner. R.R., and Patmont. C.R., *Prediction of Nuisance Periphytic Biomass: A Management Approach*, Water Research 23, 401-405 (1989).

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

ton levels in Flint Creek could be controlled, assuming they are excessive at this time. Nonetheless, the local community will be forced into major resource expenditures.

Weight of Evidence/Conditional Probability Approach

In response to extensive criticism and concern raised over the scientific basis for selecting nutrient endpoints to control nuisance algae in TMDLs, EPA developed a new approach to nutrient criteria development in 2007. This new approach can best be characterized as claiming, without specific proof, that various unrelated information supports the need for a specific nutrient endpoint while ignoring site-specific data that contradicts that claim. EPA Region III applied this new approach when it released nutrient TMDLs for five Pennsylvania watersheds in 2008. These TMDLs established phosphorus endpoints (e.g., instream nutrient criteria) using a weight-of-evidence approach that resulted in endpoints for TP ranging from 0.040 – 0.025 mg/L. These values were applied even in watersheds never listed as nutrient impaired and where site-specific data demonstrated that plant growth was not excessive. The weight-of-evidence cited by EPA in these TMDLs included distribution-based evaluations of instream nutrient concentrations, decisions of nearby states, literature review, and stressor-response analyses. Data that presented contrary conclusions were removed from the weight of evidence evaluation, such that all information presented supported the need for nutrient reduction.

Individual weight-of-evidence evaluations were prepared for three Pennsylvania ecoregions: the Allegheny Plateau Ecoregion (Sawmill Creek TMDL), the Ridge and Valley Ecoregion (Paxton Creek TMDL), and the Northern Piedmont Ecoregion (TMDLs for Goose Creek, Indian Creek, and Southampton Creek). The discussion below addresses the Northern Piedmont Ecoregion evaluation (hereafter, the *Endpoint Report*).[44] This was the first report prepared by EPA's contractor, Tetra Tech. This weight-of-evidence evaluation cited 17 lines of evidence among the three groups: distribution-based evaluations (3), stressor-response evaluations (4), and literature surveys (10) **(Table 3).** These comments are generally applicable to the other weight-of-evidence evaluations since all three were prepared by the same authors and relied on similar approaches.

One might think that a final TP endpoint, based on 17 different lines of evidence should have a reasonable amount of scientific validity and confidence it is (1) necessary to protect the environment and (2) will result in use restoration. However, a review of the various lines of evidence demonstrates this is not the case because *none* of the lines of evidence are based on a scientific demonstration of cause and effect. Merely tabulating a group of studies does not demonstrate that those studies are relevant and sufficient to establish a water quality criterion or to specify an endpoint in a TMDL that is necessary and sufficient to restore designated uses. As discussed below, most of these lines of evidence do not present cause-and-effect type relationships that link a potential endpoint with a use restoration. More importantly, the analysis acknowledges that the selected TP endpoints, ranging from 25 – 40 μg/L, will not limit plant growth. Consequently, the primary EPA objective for nutrient criteria cannot be achieved by the selected endpoints. One of the most troubling aspects of this new approach is that it purposefully ignores the available site-specific data from the various watersheds, even where such data confirm that on implementation, the TMDLs will not achieve their objectives.

Table 3
Summary of Candidate TP Endpoints
(from *Endpoint Report* Table 7)

Line	Approach – Description	Endpoint (µg/L)
	Reference	**2 – 37**
1	Reference Site 75th Percentile	16 – 17
2	All Sites 25th Percentile	17
3	Modeled Reference Expectation	2 – 37
	Stressor-Response	**36 – 64**
4	Conditional Probability – EPT taxa	38
5	Conditional Probability - % Clingers	39
6	Conditional Probability - % Urban Intolerant	64
7	Conditional Probability – Diatom TSI	36
	Other Literature	**13 – 100**
8	USEPA Recommended Regional Criteria	37
9	USEPA Regional Criteria – local data	40 – 51
10	Algal Growth Saturation	25 – 50
11	Nationwide Meta-Study TP – Chlorophyll	21 – 60
12	USGS Regional Reference Study	20
13	USGS National Nutrient Criteria Study	13 – 20
14	New England Nutrient Criteria Study	40
15	Virginia Nutrient Criteria Study	50
16	New Jersey TDI	25 – 50
17	Delaware Criteria	50 - 100

■ Detailed Evaluation of New Criteria Development Approach

The *Endpoint Report* divided the lines of evidence into three categories: reference approach, stressor-response, and literature. The ''Reference Approach'' consists of three different evaluations of regionally-specific data. The Reference Approach yields endpoints that attempt to reflect ''natural'' conditions, without demonstrating that ''natural'' conditions are required to restore designated uses. The *Endpoint Report* indicated that this line of evidence was given less weight because the distribution-based approach is *less easy to link directly to use protection, given that it is based on percentiles of a frequency distribution*.[45] In fact, the *Endpoint Report* made no attempt to relate these reference approaches to use impairment. While the *Endpoint Report* characterized these methods of evaluation as different lines of evidence, it would be more appropriate to state that the regional data set was evaluated three different ways to yield similar results. At most, this represents a single line of evidence that is unrelated to use impair-

[44] Paul, Michael J. and Lei Zheng, *Development of Nutrient Endpoints for the Northern Piedmont Ecoregion of Pennsylvania: TMDL Application* (Nov. 20, 2007), available at http://www.epa.gov/reg3wapd/tmdl/pa_tmdl/NutrientEndPoint/Nutrient_End_Point_Study_for_Pittsburgh_and_Harrisburg.pdf.

[45] *Endpoint Report* at 26.

Figure 6

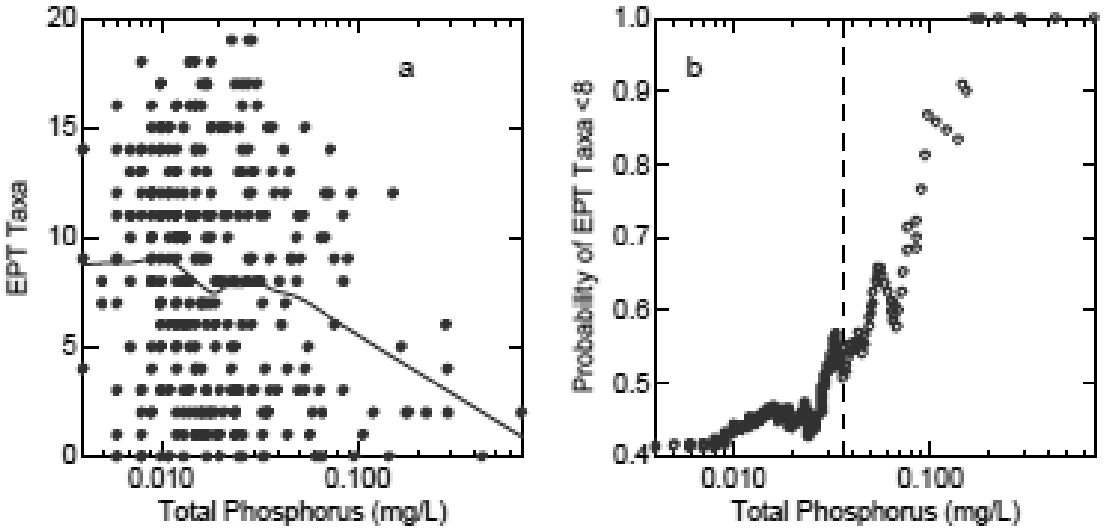


Figure 6 - Response of EPT taxa metric to the phosphorus gradient in the Northern Piedmont Ecoregion. Figure on the right shows the conditional probability of having fewer than 8 EPT taxa as TP concentrations increase.

ment or protection and provides no objective basis to conclude that achieving this nutrient level will protect stream uses.

The literature-based approach consisted of 10 different lines of evidence (*See* **Table 3**, lines 8 – 17). These lines of evidence include reference-based evaluations as well as miscellaneous reports that are unrelated to use impairments. Five lines of evidence (*See* **Table 3**, Line number 8, 9, 12, 13, and 14) apply the reference-based approach discussed above. Again, no cause and effect analysis is attempted. Algal growth saturation is acknowledged to occur when TP reaches 25 – 50 μg/L (*See* **Table 3**, line 10), the level at which the nutrient endpoints are set. Obviously, plant growth will not be limited by TP at these endpoint levels.

The last three of these reference-based approaches (Lines 12 - 14) presented an evaluation of TP data for a broad area of the United States. These three lines of evidence are based on distribution data from areas outside of the area of concern. These additional data are best described as a ''polling'' of adjacent states regarding their reference stream phosphorus concentrations. Such a survey is not a substitute for a clear scientific demonstration of cause and effect that must underlie all criteria development.

■ Stressor-Response Evaluations Flawed

The Stressor-Response approach was intended to demonstrate a cause-and-effect relationship between nutrient concentration and impairment threshold, using conditional probability analysis and change-point analysis to derive TP endpoints for three macroinvertebrate metrics and one algal metric. These metrics and the conditional probability basis for the analysis are summarized in **Table 4**.

Table 4
Stressor-Response Impairment Thresholds

Line	Metric	Impairment Threshold
4	Number of EPT Macroinvertebrates Taxa present	<8
5	Percent of Clinger Macroinvertebrates present	<52.5%
6	Percent of Urban Intolerant Macroinvertebrates	<31.5%
7	Diatom Trophic State Index	>4.5

EPT – Ephemeroptera, Plecoptera, and Trichoptera (mayflies, stoneflies, and caddisflies).

Clinger and Scrapper designations refer to behavioral and feeding characteristics.

The *Endpoint Report* indicated that these analyses were given more weight because nutrient concentrations were ''linked to specific aquatic life endpoints''.[46] This statement suggests that uses are impaired when the various metrics are not achieved. However, no evaluation was provided to demonstrate that designated uses are impaired if the macroinvertebrate or diatom metrics are not achieved. In fact, the thresholds levels identified above are not recognized by the Pa DEP as impairment levels. These values are merely the midpoints of metric scales used by the State of Maryland to characterize biological data. To our knowledge, Maryland does not use these metrics as impairment thresholds either. Merely selecting the midpoint of a data distribution provides no scientific basis to demonstrate need or reasonableness. As recently confirmed by EPA's then-existing TMDL coordinator, these nutrient endpoints might be too stringent or not stringent enough – one doesn't know.

Moreover, the report authors acknowledged that they did not determine whether TP actually was the cause of any macroinvertebrate or diatom responses being evaluated (a prerequisite for applying the conditional probability analysis) and they acknowledged that attaining the numeric instream nutrient criteria did not

[46] *Id.* at 27.

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

assure that the target metrics would be achieved (a TMDL requirement). Such a conclusion is obvious from a cursory examination of the data used to develop the endpoint. The data for EPT Taxa from the *Endpoint Report*, presented in **Figure 6**, illustrate this point.

Figure 6 shows that, even at the lowest levels of TP, the number of EPT taxa can range down to zero. At the selected ''water quality criteria,'' 0.040 mg/L, the data scatter confirms that there is no assurance that any particular EPT taxa level will be assured. This observation strongly indicates that TP is not responsible for the number of EPT taxa present. Moreover, if a stream is considered impaired if the number of EPT taxa is less than 8, it is apparent that regulating TP cannot ensure that designated uses will be restored regardless of the TP level chosen. The other lines of evidence in this group suffer similar deficiencies.

The *Endpoint Report* cited these stressor-response evaluations as the only evaluations relating nutrient concentration to environmental effect. Therefore, these evaluations should be most relevant for establishing nutrient water quality standards in accordance with the Guidelines. To the naked eye, the conclusions reached in the *Endpoint Report* as illustrated in **Figure 6** seem unbelievable.[47] However, the analysis presented to assess the stressor-response uses complex statistics to bolster the scientific validity of the proposed nutrient endpoints. Consequently, two nationally recognized experts, Dr. Domenic Di Toro[48] and Dr. Stephen Chapra[49], were asked their opinion of the new EPA analysis procedure. Excerpts of these opinions are presented below.

Dr. Dominic Di Toro (University of Delaware):

> It is not appropriate to directly compare nutrient levels in general and TP in particular to invertebrate responses. A scientifically defensible analysis must show how nutrients are affecting plant growth and then, how such plant growth is adversely impacting the ecology.
>
> ***
>
> The conditional probability approach suggested for use in Pennsylvania has, to my knowledge, never been used to derive a federal numeric water quality objective. The problem is whether the variable in question, total phosphorus concentration, is directly related to the effect that is being evaluated. Or to put it another way, is it clear that reducing the total phosphorus concentration would reduce the adverse effect, i.e., loss of benthic biota. This issue arose during the Science Advisory Board (SAB) review of the sediment criteria for which I was the Technical director on behalf of EPA. We rejected this approach in favor of a method that directly links cause to effect.
>
> ***
>
> The scientific basis for the EPA water quality criteria is one on the landmark achievements of the agency. The methodology has been adopted almost universally. It has been reviewed many times. It is based squarely on the causal relationships between the chemical being regulated and the effects being protected. The proposed nutrient criteria would be a retreat from scientifically defensible criteria to simply an expedient solution for which little or no support exists.

Dr. Steven Chapra (Tufts University):

> I have carefully read Prof. Di Toro's critique and wholeheartedly concur with all his conclusions. In particular, I strongly support his observations regarding the complexity of linking total P and invertebrate levels in streams.
>
> ***
>
> As I understand it, the proposed approach seeks to directly correlate total phosphorus (TP) concentrations with the health of invertebrate populations. Thus, phosphorus is treated as if it were a toxic substance that directly interferes with the viability and functioning of the biota. This is such a scientifically indefensible representation of the connection between nutrients and ecosystem health that I believe this its adoption would represent a grave mistake. Beyond being vulnerable to legal challenge, I am much more concerned that its adoption would ultimately be ineffective. That is, it could lead to costly controls that would not protect our precious stream ecosystems.

In response to these criticisms, EPA asked its contractor to re-evaluate the TP endpoint by assuming that the stressor-response category would not withstand scrutiny. The subsequent analysis[50] concluded that, even without the stressor-response lines of evidence, the remaining 13 lines of evidence would still support the original TP endpoint of 40 μg/L. Based on this review, EPA concluded that the criticism of the procedures is unfounded and imposing stringent TP limitations is scientifically defensible. In making this determination, EPA appears to be abandoning any need for causal relationships to set water quality standards for nutrients. Such an approach is contrary to all of the agency's prior nutrient criteria development guidance and undermines the basic purpose of the TMDL process.

Critical Evaluation of New EPA Method

■ Endpoint Derivation Flawed and Misleading

The *Guidelines* is quite clear that a simple weight-of-evidence approach is not a sufficient basis for setting a water quality criterion. Furthermore, the *Guidelines* provide that a simple regression approach between two variables (one a field response) would not suffice as a demonstration that the input variable caused the effect measured in the field. The rational behind this position is the commonly referenced scientific axiom—''correlation does not demonstrate causation.'' A re-

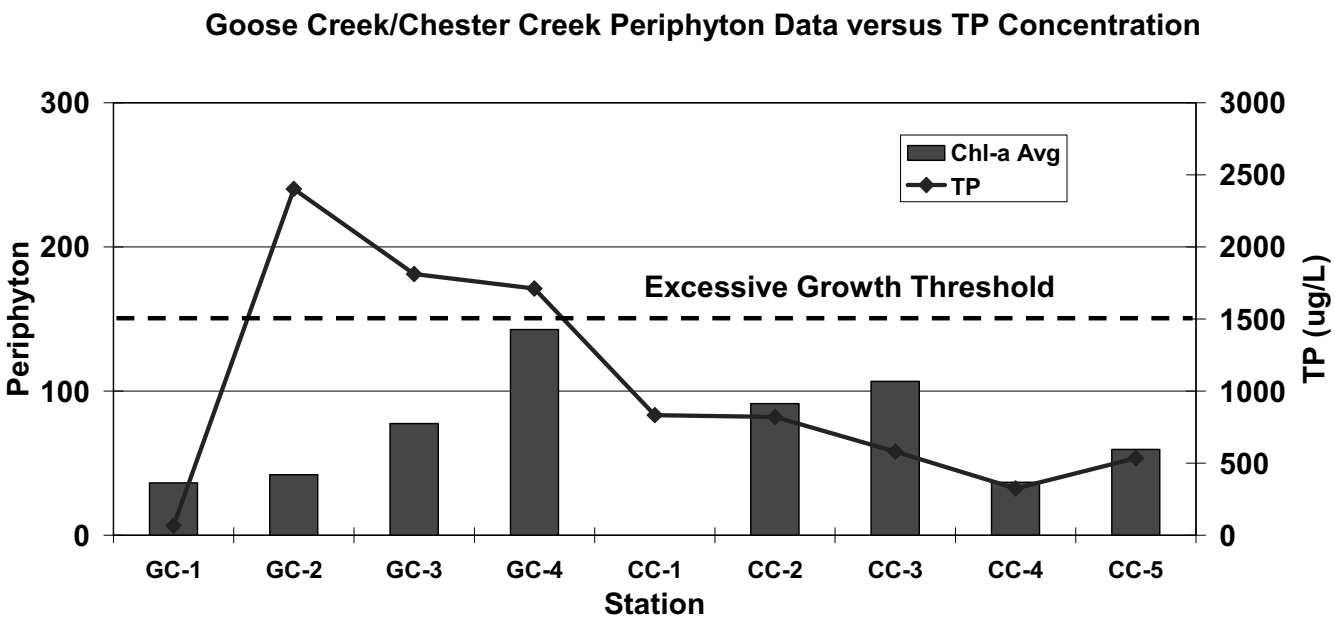
[47] It also is apparent that a few data point above 0.15 mg/L TP control and skew the regression line. Evaluation of the data below 0.1 mg/L shows there is no relationship between EPT taxa present and TP concentration as one would expect.

[48] Letter from Dr. Dominic M. Di Toro, University of Delaware, to John C. Hall, *Determining Appropriate Nutrient Reduction Requirements for Streams* (June 17, 2008).

[49] Letter from Dr. Stephen C. Chapra, Tufts University, to John C. Hall, *In Response to Inquiry Regarding the Determination of Appropriate Nutrient Reduction Requirements for Streams* (July 8, 2008).

[50] Memorandum from Michael Paul and Lei Zheng to Tom Henry, *PA TMDL Endpoints* (Nov. 10, 2008).

Figure 7



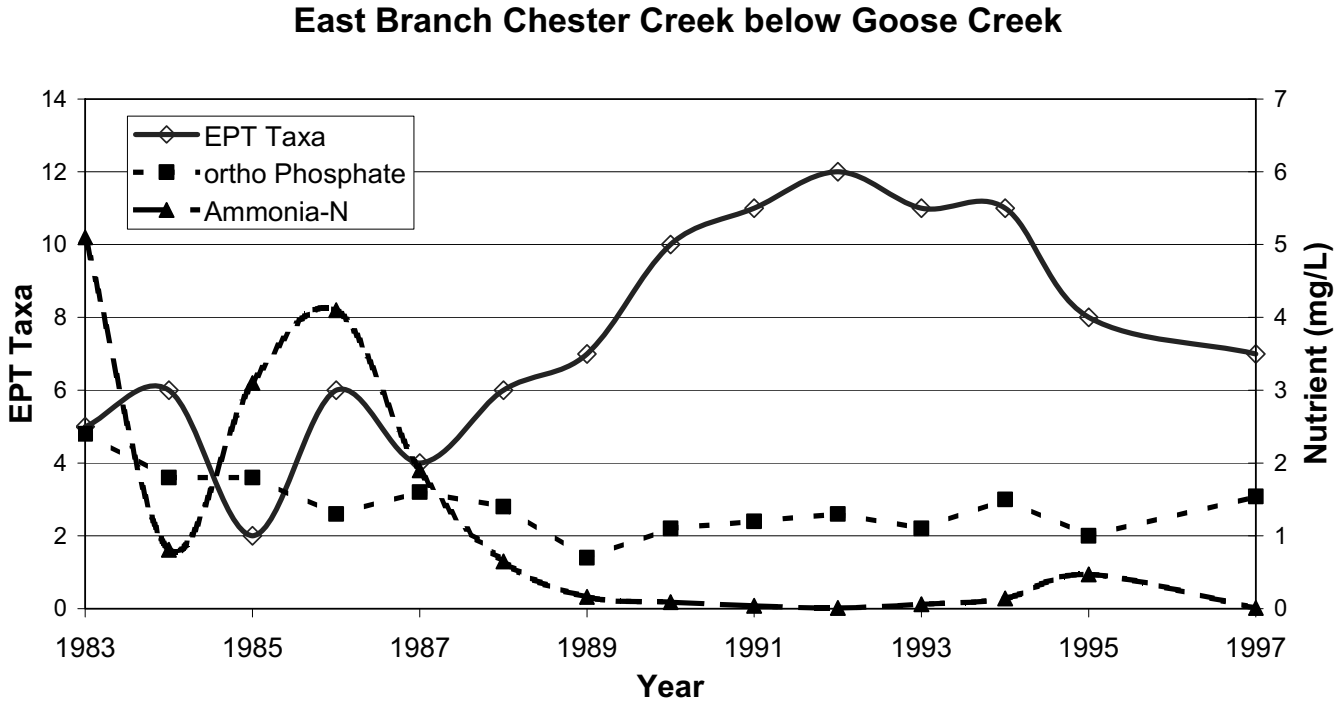
cently published EPA Office of Research and Development (ORD) paper[51] explains what is required to develop a scientifically acceptable criterion using a weight of evidence approach.

Development of numeric WQC is based on 3 basic assumptions. First, causal relationships exist between agents and environmental effects. Second, these causal relationships can be quantitatively modeled. Finally, if exposures to the causal agent remain within a range predicted by the quantitative model, unacceptable effects will not occur, and designated uses will be safeguarded. Therefore, for criteria to be valid, there must be evidence that the criteria are based on reasonably consistent and scientifically defensible causal relationships.''[52]

[51] Cormier, S.M., J.F. Paul, R.L. Spehar, P.Shaw-Allen, W.J. Berry, and G.W. Suter, *Using Field Data and Weight of Evidence to Develop Water Quality Criteria. Integrated Environmental Assessment and Management* 4(4): 490 – 504 (2008).

[52] *Id.* at 490.

Figure 8



 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

EPA's published nutrient criteria development document does not indicate that a conditional probability approach may be used for derivation of a numeric standard. This is not unexpected since that statistical method cannot provide a demonstration that regulating a pollutant at a given level provides assurances that use protection will or will not be achieved at that pollutant level. At best the method indicates the likelihood (i.e., the probability) of encountering the condition being evaluated for a given pollutant concentration used in the regression. Finally, both the *Guidelines* and the *Rivers and Streams Document* are replete with statements underscoring the need to understand the toxicology (cause and effect) of the substance. To set a numeric standard, a state must determine that the pollutant of concern is the direct cause of the adverse effect being measured. Simple regressions, conditional probability, and weight-of-evidence analyses provide no such confirmation.

Another critical flaw in the weight of evidence approach is that it fails to consider relevant site-specific information. The basic purpose of a TMDL action is to restore a documented use impairment. *EPA's Protocol for Developing Nutrient TMDLs: First Edition* states that TMDL development is site-specific.[53] The selection of the TMDL indicator (e.g., the parameter indicative of designated use attainment) is similarly site-specific. These indicators include ''both causal factor indicators (primarily, the nutrients that stimulate plant growth) and biological response indicators (which provide information concerning the impacts of nutrients on water quality). Because of the site-specific nature of TMDLs and the complexity of watershed processes, no one indicator will satisfactorily meet all of the requirements above.''[54] The *Guidelines*, the *Rivers and Streams Document*, and TMDL rules are clear that site-specific information must be considered if it shows that the suggested standard is misplaced. The new EPA approach expressly ignores such information if it contradicts the outcome of the analysis. In each case where the new standards were applied in Pennsylvania, it was acknowledged that habitat degradation (sedimentation/channelization) was the root cause of any documented changes in invertebrate populations. Site-specific regressions were provided for each watershed to demonstrate that, in fact, there was no relationship between nutrient levels and invertebrate populations in the various streams where such data were available. EPA simply chose to ignore those data and analyses, claiming the new procedures provided sufficient confirmation that nutrients were the cause of stream impairments. Rules aside, this approach is certain to squander local and state resources on ineffective pollution reduction measures.

■ TMDL Development Requirements

Adverse biological effects are caused by excessive plant growth which, it turn, is controlled by multiple physical and biological factors in addition to nutrient levels. Algal growth should be the controlling consideration with regard to developing appropriate TMDL corrective measures. In many situations, nutrients may not be the controlling factors regulating plant growth. In defense of the new approach, Tetra Tech conducted a literature review of nutrient-algal growth relationships[55]

> ''Study results summarized as part of this literature review support the assertion that while a relationship may exist between periphyton growth and nutrients, the dynamics change as a function of multiple factors. These factors include antecedent conditions, water temperature, pH, light availability, flow regime, and grazing, among others. Nutrient levels may be secondary to other determinants of biomass and growth such as light, disturbance, and grazing.''[56]

Given this dependence upon site-specific conditions, a consideration of site-specific data is mandated, but neither the weight-of-evidence approach nor the individual TMDLs considered such data. If they had, it would have been obvious that the selected endpoint would not control the biological responses.

Periphyton data provided by EPA as part of the final TMDL for Goose Creek confirmed that plant growth was rather minimal and well below the level EPA thought could cause adverse impacts (i.e., > 150 mg/m^2 as a growing season average) even though nutrient levels were an order of magnitude greater than the weight-of-evidence endpoint (**Figure 7**). This result was not surprising given that the stream tree canopy was intact. Elsewhere in neighboring watersheds, periphyton growth achieved levels in excess of 400 mg Chl-a/m^2 in response to TP < 20 μg/L where the tree canopy was largely absent (*See*, **Figure 1**). Yet, the weight of evidence approach would seek to control nutrients levels where excessive plant growth did not occur and ignore adverse impacts of excessive plant growth where nutrient levels were low.

Although the stressor-response assessments were based on invertebrate impairments, EPA ignored site-specific invertebrate data for the Goose Creek watershed, which it had referenced in the TMDL document. Those data confirmed that phosphorus levels in allegedly impaired segments of the watershed were unrelated to invertebrate populations (**Figure 8**). In fact, some of the highest phosphorus levels in the ''impaired'' segment were associated with the best invertebrate population readings.

Figure 7 confirms that nutrient control is not necessary to limit excessive plant growth in Goose Creek since plant growth is not causing or contributing to use impairments as determined by EPA itself.[57] Figure 8 convincingly demonstrates that phosphorus is not a significant stressor for EPT taxa, as the number of taxa are shown to significantly increase, apparently in response to decreases in ammonia concentration, while phosphorus concentration remains relatively constant at a level 25 times greater than the conditional probability change point calculated in the *Endpoint Report*.

For Paxton Creek, Figure 9, the impaired and unimpaired segments of the creek had similar TP levels, all well above the 25 μg/L level EPA's weight of evidence

[53] *TMDL Guidance* at 2-6.

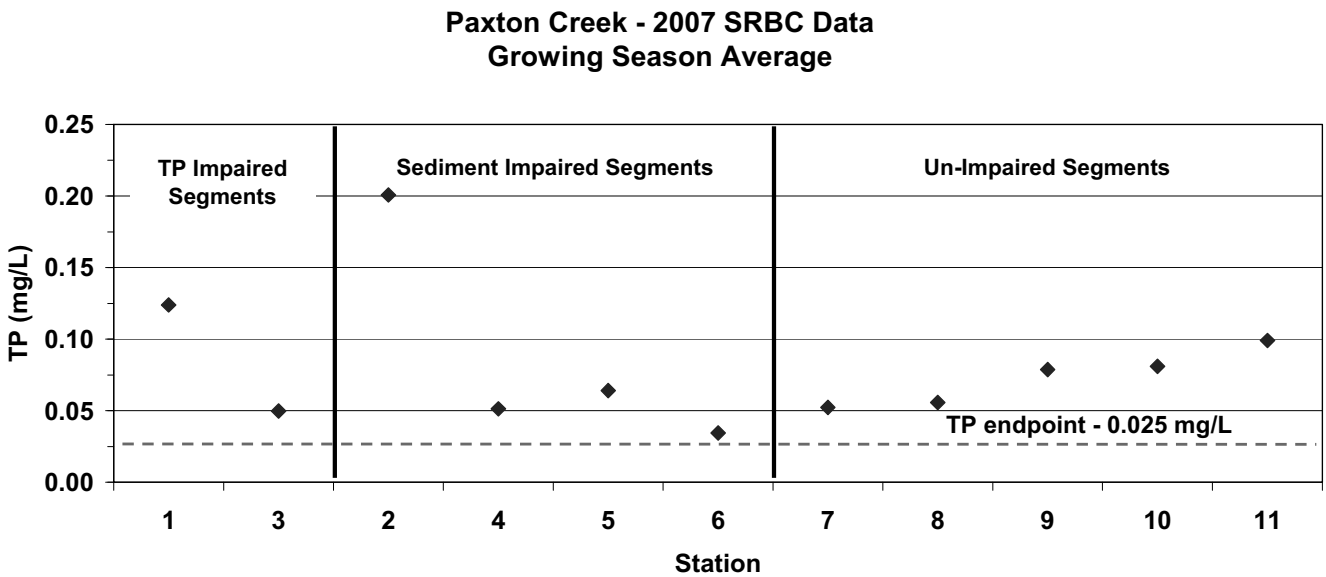
[54] *Id.* at 4-3.

[55] Tetra Tech Inc., *Literature Review to Support Selection of Nutrient TMDL Endpoints for Northern Piedmont Ecoregion Streams in Southeastern Pennsylvania* (January 2008).

[56] Tetra Tech at 18.

[57] EPA's TMDL Response to Comments stated that periphyton levels ranging from 200 – 300 mg Chl-a/m2 was considered acceptable.

Figure 9



approach claimed was necessary to ensure invertebrate protection. These data illustrate the conundrum faced by the weight of evidence method. Stream segments, evaluated as unimpaired using site-specific measures of biological metrics as the basis for evaluation, would be assumed to be impaired for these same metrics based on exceedance of the weight of evidence endpoint. It should be obvious that such an approach is misguided. In actuality, segments exceeding the endpoint concentration were specifically determined to be unimpaired, demonstrating that the weight of evidence endpoint is an unreliable indicator of impairment.

SAB Peer Review Requested

The new weight-of-evidence approach to derive nutrient water quality criteria represents a radical departure from the Agency's historical guidance on criteria development and nutrient TMDL implementation. It is apparent that the new approach is contrary to a series of ''bedrock'' scientific principles for criteria development that the Office of Water has relied upon for decades, including:

- Numeric criteria must be based on documented dose/response relationships between the pollutant and a use impairment (versus assuming the pollutant is causing the problem and ignoring data to the contrary)
- Numeric standards must be set at the level found both necessary and sufficient to protect uses (versus setting the standard where the probability of impacts is decreased even if the stressor response is extremely weak)
- Nutrients are not directly toxic to invertebrates but affect plant growth (versus ignoring the degree of plant growth occurring and assuming that nutrients directly impact invertebrate populations)
- Confounded data may not be used to develop a numeric standard (versus assuming all measured field responses are due to a pollutant, even where the data show this is not true), and
- Site-specific data, when available, must be considered in determining whether a numeric standard is necessary and will achieve its intended level of protection (versus ignoring the site-specific data and assuming that the generalized conditional probability analysis justifies pollutant restrictions).

EPA's decision to abandon well established scientific principles and requirements and alter its published criteria development approaches is a major federal action of national importance. Such regulatory actions must be transparent and undergo detailed scientific assessment. Federal peer review procedures require that new, innovative or controversial scientific procedures used to establish regulatory program requirements must first undergo peer review before they are used in a regulatory context. On Dec. 16, 2004, the Office of Management and Budget (OMB) issued a final bulletin to all agencies establishing that influential scientific information shall be peer reviewed before it is disseminated by the Federal government [58] and EPA updated its own peer review policy to accommodate the OMB requirements (EPA/100/B-06/002, May 2006). Although agencies have discretion to choose the specific type of peer review to employ, the duty to conduct a peer review is not discretionary.[59] In determining the extent of the peer review necessary, the OMB bulletin stated that ''[m]ore rigorous peer review is necessary for information that is based on ***novel methods*** or presents complex challenges for interpretation. Furthermore, the need for rigorous peer review is greater when the information contains ***precedent-setting*** methods or models, presents conclusions that are likely to ***change prevailing practices***, or is likely to affect ***policy decisions*** that have a significant impact.''[60] There is no serious question that EPA's attempt to use a new scientific approach to nutrient criteria derivation, at odds with its published scientific approach, meets every component of the OMB Bulletin justifying a detailed peer review.

Given these concerns, the agency was petitioned to conduct a full SAB peer review on the use of conditional

[58] 70 *Fed. Reg.* 2664 (Jan. 14, 2005)
[59] *Id.* at 2675.
[60] *Id.* at 2668 (emphasis added).

 COPYRIGHT © 2009 BY THE BUREAU OF NATIONAL AFFAIRS, INC. ISSN 0013-9211

probability and weight of evidence, as presented in the *Endpoint Report*, to derive nutrient water quality criteria[61]. EPA responded, reaffirming its position that the nutrient endpoints developed using conditional probability-based and other lines of evidence are considered technically and legally sound. [62] Nevertheless, EPA agreed to peer review this approach in the context of providing broader national guidance. On April 27, 2009, a Federal Register notice (74 *Fed. Reg.* 19,084) requested nominations of experts for the review of technical guidance on nutrient criteria development.

It is clear that EPA intends to promote the weight-of-evidence approach nationwide for development of nutrient water quality standards. Thus, all dischargers potentially face stringent nutrient endpoints developed in the manner described above, regardless of actual environmental need or efficacy. Before that happens, however, the weight-of-evidence approach will undergo a full SAB peer review. This review is open to the public and will allow for input from interested stakeholders as part of the overall peer review process. Given the dubious linkage between the endpoints derived and use restoration (in addition to the significant detrimental economic/energy impacts associated with misplaced standards), all parties should voice their concerns to the SAB.

Summary and Conclusions

The weight of evidence approach to developing numeric nutrient water quality standards, as currently espoused by EPA, is seriously flawed. The approach ignores the basic foundation behind all other water quality criteria and TMDL decisions– cause and effect. Even if the weight of evidence approach could overcome this flaw, the resulting endpoint should not be used if site-specific data confirm that the selected endpoint is not applicable to a given stream. However, as illustrated in the case of Goose Creek, the agency intends to ignore site-specific information to the contrary and apply ''weight of evidence'' endpoints without regard to actual need. Finally, the ''weight of evidence'' approach presents a single-minded focus on nutrient control rather than considering the multitude of physical/biological factors that influence ecological response. A large body of evidence is available confirming that these factors may provide the key to use protection. To ignore these factors and force nutrient reductions that would not otherwise be necessary to protect stream uses does not make good sense, either environmentally or economically.

Notwithstanding these concerns, EPA has set its course. The regulated community has an opportunity to redirect EPA's program by participating in the SAB peer review.

[61] Letter from John C. Hall to Stephen L. Johnson, EPA Administrator, *Request for Peer Review of New EPA Region III Approach to Developing Instream Standards for Nutrients* (Aug. 21, 2008).

[62] Letter from Ephraim King to John Hall (Dec. 24, 2008).