A Synopsis on

Developing Numeric Nutrient Criteria for Freshwaters

JoAnn M. Burkholder, Ph.D.

2018

Executive Summary

Nutrient pollution (“cultural eutrophication”) is among the most important sources of impairment to U.S. surface waters. Numeric nutrient criteria are needed to protect the nation’s surface waters because they create clear, direct goals aimed at the two key pollution parameters (total nitrogen, TN, and total phosphorus, TP) that repeatedly have caused undesirable impacts on aquatic life, drinking water, and recreation.

Water quality criteria can be either narrative (for example, “waters of the state shall be free from unnatural plant or algal growth”) or numeric (for example, Wisconsin Rivers shall not exceed a concentration of 0.1 mg/L total phosphorus). The United States Environmental Protection Agency (U.S. EPA) has long recognized that numeric nutrient criteria are needed to control nutrient pollution. This paper first summarizes federal regulations, recommendations, and guidelines related to numeric nutrient criteria, followed by a synopsis of the science background about nutrient issues in surface freshwaters, including definitions of basic terms and descriptions of important concepts.

Numeric criteria are considered here both for the two main forms of nutrient pollution (or "stressors"), nitrogen and phosphorus, and for “response” parameters that can be measured to determine the extent to which nutrient pollution has already harmed waterbodies. While the floating (suspended or sestonic) microalgae in lakes, reservoirs, and larger rivers, or the benthic algae growing on the bottom of the waterbody in smaller streams, are usually the only responders measured in assessing impacts of nutrient pollution for criteria development, there actually are four key-response groups of primary producers – (i) suspended microalgae, (ii) benthic microalgae and macroalgae, (iii) floating macroalgae such as filamentous (string-like) green algae, and (iv) vascular plants. The response of each of these potential responder groups to nutrient pollution is described.

Five available approaches for developing numeric nutrient criteria, suggested by the U.S. EPA, fall within two major lines of reasoning. The first line of reasoning seeks to set criteria based on what was the natural, pre-development or “minimally impacted” levels of nutrients. The underlying goal is to maintain or restore conditions as they were historically, insofar as possible, to avoid damage to beneficial aquatic life or for other designated uses such as recreation. The second line of reasoning sets criteria based on studies of how nutrient pollution harms the designated uses of waterbodies (e.g., aquatic life support, fishing, and/or recreational uses such as swimming, drinking source-water, etc.).

(1) The reference approach falls under the “minimally impacted” line of reasoning, wherein numeric criteria are developed by estimating the natural or least altered levels of nutrient stressors and response variables. The criteria are based on the premise that nutrient levels and responses should be similar to the natural condition. (2) The literature approach, which can be under either line of reasoning, relies on suggested boundaries for enrichment classification of lakes and reservoirs, and of rivers and streams. (3) The stressor/response-based modeling approach uses available field and experimental data to estimate statistical relationships between nutrient concentrations and response measures that are relevant to the designated uses. (4) The

mechanistic modeling approach develops mathematical representations of the ecological system(s), processes, and parameters (factors or variables), using equations that can be calibrated using site-specific data. Finally, (5) “hybrid” analyses include more than one approach, such as the reference (minimally impacted) condition, “weight of evidence” of different statistical assessments of stressor-response relationships, and mechanistic models.

Both lines of reasoning have strengths as well as deficiencies. The reference condition approach is vulnerable to the criticism that it may be too strict - that is, the numeric nutrient criteria selected using that approach, by seeking to preserve or restore water quality close to natural nutrient levels, may be lower than necessary to protect the designated uses of the waterbody(s) of interest. On the other hand, focus upon impacts caused by nutrient pollution could allow degradation of higher-quality waters. Consider, for example, a waterbody of interest that, in its natural condition, did not support largemouth bass. Various sportfish species thrive in higher nutrient regimes with abundant algae. If the waterbody was altered so that it supported the designated use for largemouth bass, then that alteration would involve increasing human-related nutrient enrichment, which could adversely affect naturally occurring fish species that are more sensitive to nutrient pollution than largemouth bass.

The five approaches suggested by the U.S. EPA for use in developing numeric nutrient criteria are described and illustrated using various examples. Whatever approach is used, the approach selected largely must depend on the data available for the waterbody(s) of interest (Dodds and Welch 2000). The datasets used to develop numeric nutrient criteria can be compared to the data in hand for the system(s) to assess whether the data are sufficient for use with an approach under consideration. If sufficient water quality (TN, TP, chlorophyll a) data are available for waterbody(s) of interest (e.g., data collected at least monthly during the summer growing season, over the past 5-10 years, for nutrients and chlorophyll), then numeric nutrient criteria should be developed based on localized conditions following accepted science-based procedures. This basically describes the reference condition approach using nutrient sub-ecoregions. If a strong dataset is also available for sensitive macroinvertebrate and fish biota, then stressor-response relationships can also be developed, supported by peer-reviewed literature.

It is recommended that at least three approaches should be applied as an internal check. The resulting numeric criteria indicated by two or more approaches should be similar. Considering the uncertainty inherent in understanding the complex relationships between nutrients and various response variables, the lower, more protective nutrient criteria should be used.

Other basic guidance in developing numeric nutrient criteria is included:

- Considering data needs, the criteria should be based on high-quality data, collected with sufficient frequency (recommended to be at least monthly) during the appropriate period, for example, during the summer season if the user seeks to minimize noxious summer cyanobacteria blooms. Assessment of average nutrient concentrations (TP, TN) within that period should encompass baseflow, stormflow, and wetflow conditions. Samples should be analyzed by a state-certified laboratory using strict quality control/quality assurance protocols. If a state has only sparse or dated information on a waterbody(s) of interest, data should be considered from other states for similar waters within the same ecoregion or ecosystem conditions.

- True reference or minimally impacted conditions should be used to develop criteria that are based on the reference approach.

- The numeric criteria should be protective rather than reactive. Criteria should be developed for TN and TP, the causative factors, rather than only for response variables such as algal biomass indicated by the response variable chlorophyll a, because chlorophyll a and other response variables are often poor surrogates for the total impacts of TN and TP. Reliance upon response variables alone will not protect waters because by the time they indicate danger, the waterbody can be so damaged that it can take years or decades to recover. Also, criteria for stressors, (generally TN and TP) should be set low enough that they will prevent damage before it develops.

- Effects-based approaches should rely on protective statistical procedures, such as the first threshold identified from additive quantile regression with smoothing, that is, the lowest nutrient concentration wherein a small increase corresponds to a relatively large adverse change in metrics for desirable biota, such as the number of pollution-sensitive macroinvertebrate species. When a clear threshold is not evident and/or when the desirable metric gradually changes as nutrient pollution increases, then a line-drawing procedure needs to be used, as adopted by the U.S. EPA in setting other protective numeric criteria (Stephan et al. 1985).

- If a state agency includes response variable(s) in its numeric criteria, the most abundant, noxious key responder primary producers in the waterbody(s) of interest should be targeted.

- Arithmetic means should be used to set numeric nutrient criteria rather than geometric means or medians. Geometric means are biased low, and do not reliably estimate average water quality conditions. Arithmetic means are unbiased, easier to calculate and understand, and more protective.

- Compliance should be based on independent consideration of TN, TP, and response variables. Each should "stand alone" to protect waterbodies from nutrient pollution-related degradation. That is, exceedance of all or some grouping of these variables should not be required before a waterbody is assessed to be in violation of the numeric criteria. Rather, if any one of the causal/response variables is exceeded, then the waterbody should be characterized as in violation. For example, a violation should not require exceedance of both TN and phytoplankton chlorophyll a, or both TP and phytoplankton chlorophyll a. Independent consideration of these criteria is needed because nutrient pollution can adversely affect beneficial aquatic life through means other than increased phytoplankton growth, such as stimulation of other key responder groups (benthic algae, floating macroalgae, or vascular plants). Sampling locations should be carefully selected for compliance assessment, and locations where measurements are taken should be specified.

A final section of this white paper goes into more detail on "doing it wrong," that is, non-protective actions that should be avoided in developing numeric nutrient criteria:

- Nutrient criteria should not be set reactively, that is, when pollution-tolerant species predominate and pollution-sensitive species have been nearly or totally eliminated.

- Waterbodies that have been substantially compromised by nutrient pollution should not be used as "minimally impacted."

- Data from true minimally impacted waters in the same ecoregion should be used rather than rejected, even if they are from similar waters in other states, in setting numeric nutrient criteria if such waters are no longer available in the state setting criteria.

- "Add-on," concocted procedures should not be used that defeat the purpose of the accepted technique being invoked.

- Response variables such as suspended or benthic algal biomass (for example, considered as the indicator, chlorophyll a concentration) should not be used as a "surrogate" for numeric TN and TP criteria, because TN and TP strongly influence some water quality conditions that are not related, or only weakly related, to algal chlorophyll levels.

"Doing it wrong" also pertains to models that are wrongly used to develop non-protective numeric nutrient and response variable criteria. The key to model assessment lies in determining the major underlying assumptions used to build the model and the quality of the dataset used to construct it. Models should not be based on few or no actual data from the waterbody(s) of interest, and they should not be constructed using erroneous assumptions and other non-science-based information.

I. Introduction

Nutrient pollution (also referred to as "cultural eutrophication"), especially by two nutrients, phosphorus (P) and nitrogen (N), is among the most important sources of impairment to our nation's surface waters (National Research Council 2000). To address this major nationwide problem, EPA has long recommended that states adopt numeric criteria for total nitrogen (TN) and total phosphorus (TP) (U.S. EPA 2010). The purpose of this paper is to describe how that can best be done.

Two measures, concentrations and loadings, are frequently used to describe nutrient conditions in surface waters. Concentration refers to the amount (mass) of nutrient in a defined volume of water (such as micrograms of ammonium per liter of water). Load refers to the total amount of nutrient entering the water over a specific period of time (such as kilograms of nitrogen per year). Load is calculated as nutrient concentration x volume (such as nutrient concentration x volume in effluent discharged by a point source). Ambient nutrient concentrations are analogous to a "snapshot;" they indicate the supply level, when the sample was collected, that was available to support microbial (algal, bacterial) production (Hecky and Kilham 1988, Mallin et al. 2006). In contrast, nutrient loadings assess the overall nutrients supplied to an aquatic ecosystem over time, generally on a per-year basis (National Research Council 2000).

The U.S. Congress directed states to establish water quality standards that "consist of the designated uses of the navigable waters involved and the water quality criteria for such waters based upon such uses" (33 U.S.C. § 1313(c)(2)(A), 40 C.F.R. § 131.2). EPA regulations specify that "[s]uch criteria must be based on sound scientific rationale and must contain sufficient parameters or constituents to protect the designated use. For waters with multiple use designations, the criteria shall support the most sensitive use." 40 C.F.R. § 131.11(a)(1). A water quality standard defines the water quality goals for a waterbody, or portion thereof, by designating the use(s) to be made of the water and by setting criteria necessary to protect the uses (Clean Water Act, 40 CFR § 131.2).

Thus, numeric nutrient criteria should be designed to protect the designated uses of the waterbody of interest (see Section IV below). Water quality criteria generally are intended to measure the ambient quality of a waterbody, and for most pollutants including nutrients, the biological responses of aquatic life are triggered by the concentration in the water. If the concentration criteria are exceeded, then loading thresholds for the waterbody, or total maximum daily loads (TMDLs), may be developed.

The U.S. EPA (2000a) recommended development of numeric N and P concentration criteria, as well as numeric criteria for response variables such as suspended microalgal biomass as chlorophyll a, for U.S. states as an important step in controlling nutrient pollution to the nation's surface waters. Two other important considerations for control are N:P ratios and N and P loads, as explained in Section III below, but the main focus of this paper is N and P concentration criteria. Guidance is provided here about how states can develop appropriate numeric N and P criteria for surface freshwaters. Numeric N and P criteria are needed in most states. In the Mississippi River watershed, for example, two states, Minnesota and Wisconsin, have adopted numeric P criteria for their lakes, reservoirs, rivers and streams, excluding ephemeral waters; Illinois has adopted a numeric P criterion for lakes

and reservoirs larger than ~8.1 hectares (20 acres) in area. None of the states in the Mississippi River watershed have numeric N criteria, except for a drinking water nitrate standard of 10 mg/L (maximum contaminant level) in accord with the federal standard (U.S. EPA rules at 40 C.F.R. § 141).

This white paper is structured to include:

- A brief summary of federal regulations, recommendations, and guidelines;
- Science background about nutrient issues in surface waters;
- Approaches for developing science-based numeric N and P criteria, along with basic guidance and supporting rationale, and an appendix for more about statistical threshold techniques; and
- "Doing it wrong" - pitfalls and precautions in developing science-based numeric criteria.

II. U.S. EPA Recommendations for Numeric Nutrient Criteria

The U.S. EPA (2017) has informed the general citizenry that "Nutrient pollution is one of America's most widespread, costly and challenging environmental problems, and is caused by excess nitrogen and phosphorus in the air and water". The agency described strengthening water quality standards, including adoption of numeric nutrient criteria, as an important tool for meeting clean water goals, because "numeric criteria provide measurable water quality-based goals that are easier to implement" [and clearer to enforce] than narrative criteria (U.S. EPA 2016a).

The U.S. EPA (1998, p. iv) directed states to adopt ambient numeric nutrient criteria (measurable benchmarks) for streams and rivers, including stipulated concentrations, duration, and frequency. To facilitate that effort, the agency provided a series of nutrient criteria guidance documents and recommended criteria for nutrients (causal variables TP and TN or N forms such as nitrate) in lakes and streams within designated nutrient ecoregions (also see Section IV below). The U.S. EPA developed 14 major ecoregions for the nation based on present land use, and various sub-ecoregions within each ecoregion. The recommended criteria for each were based on minimally impacted conditions, defined as having natural background nutrient concentrations, unaffected (true reference conditions) or minimally affected by nutrients from human activities. The targeted numeric nutrient criterion was defined as the 75th percentile of true reference stream data (past ~decade) or, if reference streams could not be found, the 25th percentile of data for all streams within a nutrient sub-ecoregion. States were given the option of using the numeric nutrient criteria recommended by the U.S. EPA, or developing their own scientifically defensible numeric criteria that protect the designated uses of surface waters (Section IV).

The U.S. EPA (2000a, 2001a) alternatively recommended other approaches such as use of literature values under conditions when data are sparse; or use of stressor-response relationships to derive numeric nutrient criteria if sufficient data are available, mainly by identifying threshold or changepoint concentrations at which large changes in biological metrics (e.g. algal biomass as chlorophyll a, or sensitive biota such as aquatic macroinvertebrate insects) occur as a result of increasing nutrient concentrations; or use of a

combination of approaches (see Section IV). The U.S. EPA guidance explained that biological thresholds are valuable for setting nutrient criteria because there is a direct link between biological responses and protection of designated uses for aquatic life. The U.S. EPA (2016a) maintains a Nutrient Scientific Technical Exchange Partnership and Support Program (N-STEPS) that provides technical assistance to states for developing numeric nutrient criteria.

The U.S. EPA has contributed other guidance on nutrient management, such as plans directed at certain waterbodies or watersheds. For example, the federal Gulf Hypoxia Action Plan calls for each state in the Mississippi River basin to produce a plan to reduce the amount of N and P carried from rivers in the state to the Gulf of Mexico. While many states lie in the Mississippi River drainage, six of them - Iowa, Illinois, Missouri, Kentucky, Tennessee, and Arkansas - are the largest contributors to the nutrient supplies that fuel the Dead Zone hypoxia in the Gulf of Mexico (Figure 1).

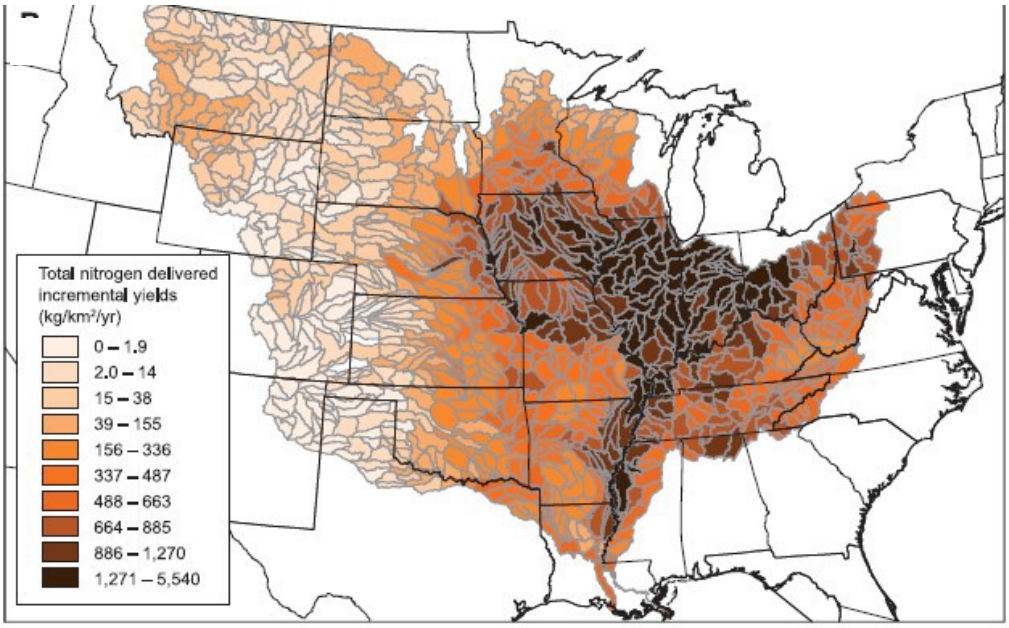
III. Science Background

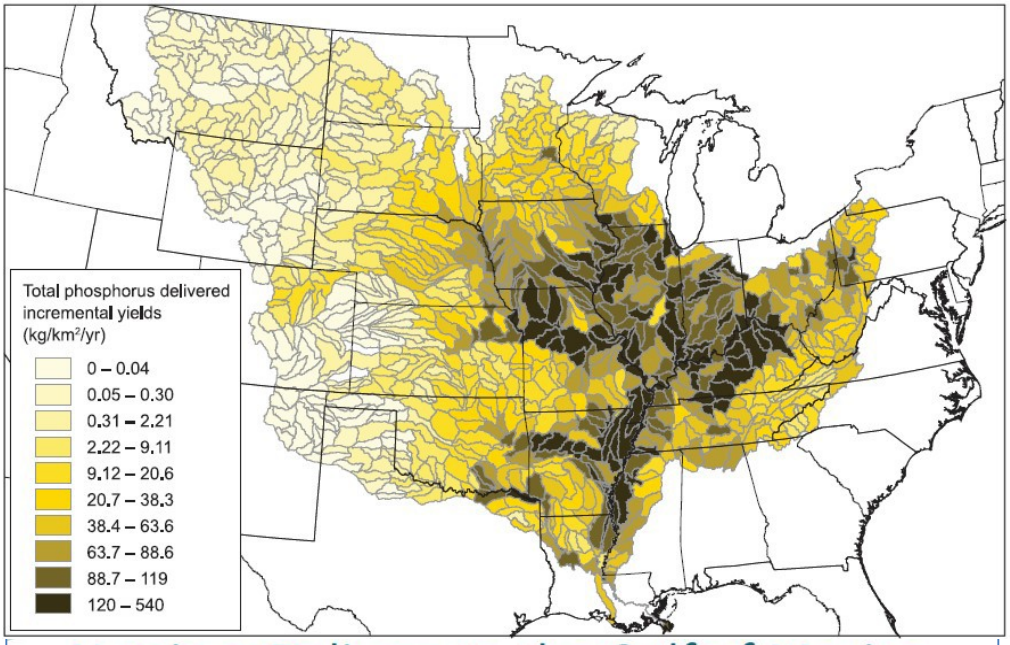
III.a. N and P Pollution, a Ubiquitous Problem in Surface Waters

Cultural eutrophication is the process of N and P contamination of surface waters from sewage, urban runoff, cropland and livestock agriculture, and air pollution, which add high levels of N and P, relative to what is natural, so that the affected aquatic ecosystem is pushed out of balance to an unhealthy state (Figure 2). Human-derived sources of nutrient pollution have rapidly changed water quality and aquatic community structure within the past ~250 years of industrialization and rapid population growth (Cloern 2001, Wetzel 2001). Throughout most of the world increased nutrient enrichment (or nutrient pollution), is a major issue (Burkholder and Glibert 2013).

Despite the advances in treatment of human sewage during the late twentieth century, limits or bans on use of phosphorus for certain domestic and industrial practices, declines in use of agricultural fertilizers in some regions, and modest improvements in environmental education in localized areas, the ongoing increase in human population growth continues to increase nutrient loadings to aquatic ecosystems (Burkholder and Glibert 2013). Cultural eutrophication promotes major adverse shifts in the structure of algal/plant and animal communities, generally affecting dominant components of every trophic level from microbial decomposers to the larger animals at the top of the aquatic food web, and significantly reducing biodiversity.

Why should N and P be the focus for numeric nutrient criteria? Because, relative to the available environmental nutrient supplies needed for algal and plant growth under natural conditions, N and P are the essential nutrients that algae and plants run out of first (Vallentyne 1974, Wetzel 2001). So, under natural conditions, it is the lack of these nutrients that prevents excessive plant or algal growth. Total nutrient levels (both N and P) have been more strongly correlated with suspended microalgal biomass than soluble nutrient forms (e.g, Dodds et al. 1997). Both nutrients are important to control because in surface waters, N and P together describe algal biomass better than either nutrient alone (Smith 1982, Prairie et al. 1989, Dodds and Smith 2016); “more P means more chlorophyll with per unit N, and vice versa” (Dodds and Smith 2016).





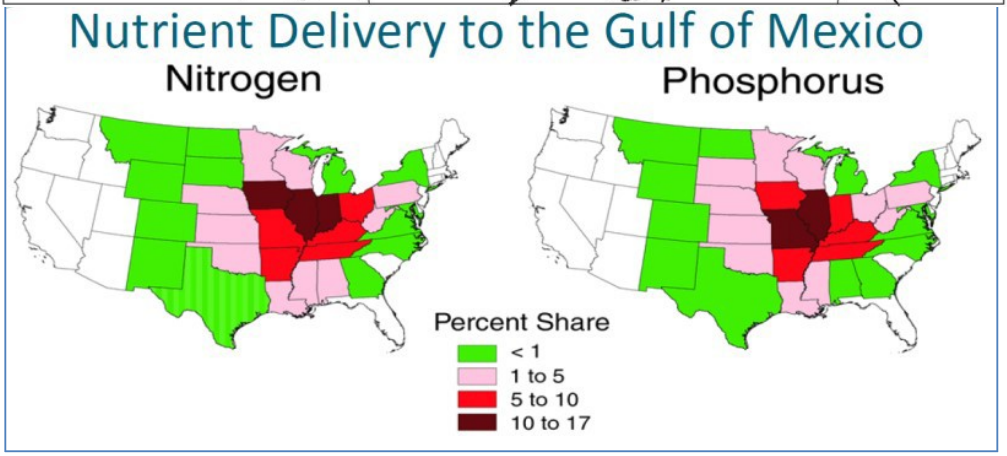


Figure 1. N and P delivery to the Gulf of Mexico Dead Zone. Upper, middle panels - from Robertson et al. (2009); lower panel from Manner (2014); also see Alexander et al. (2008).

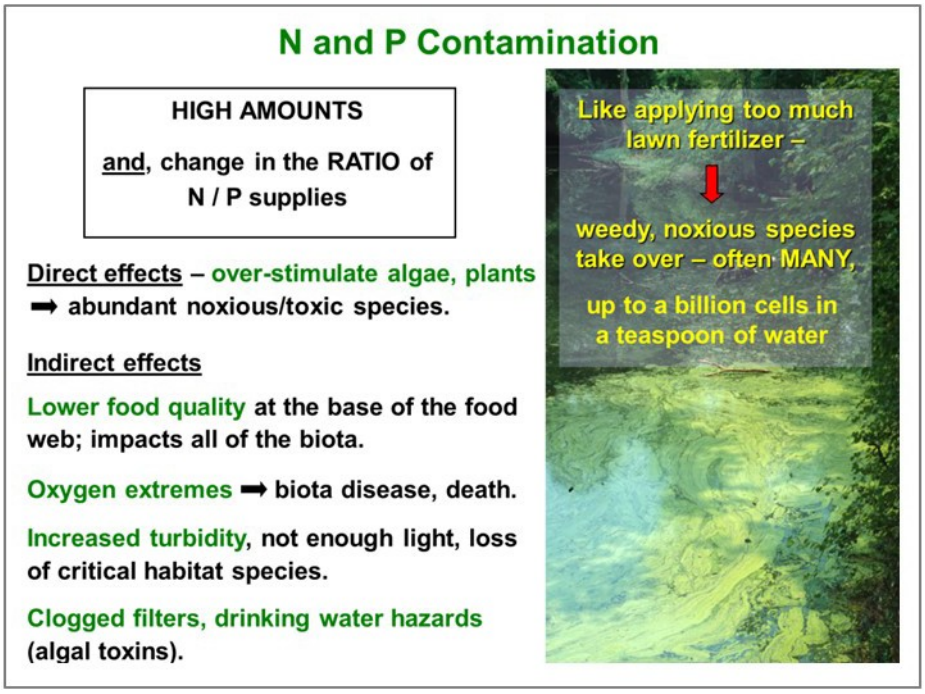


Figure 2. Brief summary of major impacts of nutrient pollution in surface waters.

When an essential nutrient is needed for the organism to complete its life history and it is not substitutable, supply of that nutrient has a direct effect(s) on the organism's physiology and growth. No other nutrient can be substituted for N or P in meeting an organism's physiological needs. The scientific recognition of "primary nutrient limitation" dates back to Liebig's Law of the Minimum in the 1840s, from agricultural literature on crop yield (Liebig 1847): If a nutrient is limiting, then its amount in algal or plant cells will decline to a low level, and the nutrient in lowest concentration relative to the specific requirements for that alga or plant will limit growth. That is, the nutrient in least supply relative to the need of the plant or alga will limit growth, whereas increases in plentiful nutrients will have little effect.

Nitrogen is an essential nutrient, required for organisms to make amino acids, proteins, enzymes and coenzymes, nucleotides, and nucleic acids (genetic material), chlorophyll pigments, and various other essential algal pigments. The forms of N that many algae and plants generally use for growth are ammonia (ionized form, ammonium) and nitrate, two types of inorganic N. Algae and plants can also use certain forms of organic N for growth, such as urea and some simple amino acids (Hecky and Kilham 1988). In natural (background) conditions, most of the TN consists of organic N; inorganic N forms are usually low, typically < 200 µg NO_3^-N/L and < 20 µg NH_4^+N/L, respectively (Stanley and Maxted 2008, and references therein). In assessing aquatic systems for N abundance to primary producers, either TN or inorganic N forms are usually considered. Processes involving N in surface waters are more complicated than those involving P, partly because nitrate, another inorganic N form called nitrite, and ammonia/ammonium can be toxic to algae and plants, depending on the concentration; low amounts are beneficial, but high amounts are detrimental and these nutrient forms act as chemical poisons. Moreover, ammonia/ammonium can be an oxygen-demanding substance, because it takes up oxygen

during the process (called nitrification) when much of it is converted to nitrate (Stumm and Morgan 1995).

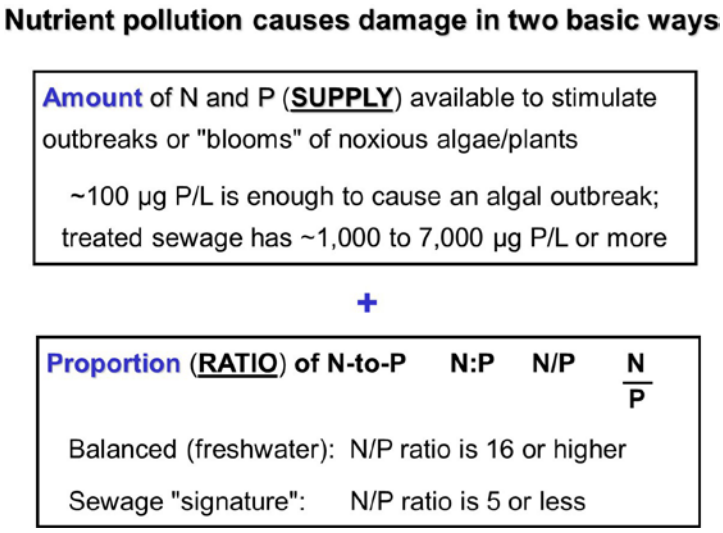


Figure 3. Chart depicting the importance of BOTH nutrient supplies and nutrient supply ratios in cultural eutrophication. Note that these atomic (molar) ratio numbers typically refer to conditions in the summer growing season.

Phosphorus is also an essential nutrient; it is required for organisms to make the energy currency for cells, ATP, and other nucleotides; P is also a component of nucleic acids such as DNA, certain proteins, several important coenzymes, membrane phospholipids that are required for cell survival; and P is attached to many different substances that are important in photosynthesis and respiration. Algae and plants generally use the inorganic phosphate ion (sometimes referred to as orthophosphate, or soluble reactive phosphorus - SRP, or $PO_4^{-3}P$) for growth. In natural (background) conditions, inorganic P is usually very low in supply, typically < 5-10 µg $PO_4^{-3}P$ (Wetzel 2001). If surface waters have phosphate concentrations exceeding 25 µg/L (25 parts per billion), that is an indication of nutrient pollution from sewage (treated effluent P is mostly phosphate), animal wastes (which can also be high in phosphate), mining sources, fertilizer, or other sources (Carpenter et al. 1998, Correll 1998). Algae and plants (as well as bacteria) luxury-consume phosphate (that is, they take much more phosphate up than they need at the time), so much so that the phosphate ions left in solution are usually much, much lower in supply than the P stored in their cells. For that reason, in assessing aquatic systems for P abundance to primary producers, the TP concentration, which includes the P in the algal and plant cells, is usually used (Wetzel 2001).

Nutrient pollution can damage aquatic ecosystems in two basic ways: First, through an increase in the available amounts (supplies) of N and P that stimulate outbreaks (blooms) of noxious algae and plants. About 100 µg P/L (or even less) commonly is enough to cause an algal bloom in lakes and reservoirs, if other conditions (temperature, light, etc.) are conducive, but partially treated sewage has ~1,000 to 7,000 µg P/L; and second, through a shift in the proportion (ratio) of N relative to P supplies (Figure 3, and below). Surface waters affected by nutrient pollution usually have both problems: The N and P supplies are extremely high in comparison to natural ("background") conditions that historically occurred, and the N:P ratio is skewed so that the system has been pushed into what is referred to as "stoichiometric imbalance" (Burkholder and Glibert 2013).

The present status of many U.S. surface waters, including those in the Mississippi River basin, is that they are not only over-enriched with P and N but also in a trophic state of stoichiometric imbalance, which is worse than simply eutrophic (nutrient-rich). Stoichiometric imbalance is defined as a forced trophic state in a waterbody that develops when the supply of one nutrient (generally P or N), is altered either due to enrichment from

human activities or management-related nutrient control (Burkholder and Glibert 2013). A common "signature" N:P ratio for raw or partially treated sewage is 5 or less (Metcalf and Eddy 1991, Henze et al. 1997). Sewage treatment historically (and often still) has targeted P removal without N removal (Glibert et al. 2011).

III.b. Impacts of Nutrient Pollution on Surface Waters

The most basic impacts from nutrient pollution on surface waters (Figure 2) are discussed below. The science of nutrient pollution, like any other discipline, has its own language. The good news is that only a little more than a dozen terms is needed to "speak this language fluently." The following information is taken from Burkholder and Glibert (2013) and Burkholder (2014) unless otherwise noted.

III.b.1. Key Responders of aquatic food webs in response to nutrient pollution, in conditions where light is abundant, are the primary producers (algae and plants). It should be noted that in low-light conditions such as blackwater or heavily shaded streams, bacteria and animal-like microbes (protozoans) can instead be the key responders to nutrient pollution (e.g., Mallin et al. 2006) because they do not need light for their metabolism. This paper emphasizes primary producers as the major responders to nutrient pollution. They include algae and plants that use water and sunlight to "fix" carbon dioxide from the air into organic carbon in their cells. This type of photosynthesis is sometimes called "oxygenic" because oxygen is produced during this process. All algae and plants with oxygenic photosynthesis have the "universal plant pigment," chlorophyll a, which is essential for that process.

Four general groups of aquatic primary producers, based on appearance and habitat, are most responsive to nutrient pollution. They include suspended (sestonic) microalgae, floating filamentous macroalgae, benthic (attached to the stream bottom mud, rocks, or other substrata) microalgae and macroalgae, and plants (Figure 4), described below. It is important to note that, of the four groups, data are usually available only for floating microalgae. Floating filamentous algae and floating plants are very difficult to sample quantitatively for biomass, which is usually taken as dry weight per unit lake or stream area, or simple percent cover of an area of interest. Attached algae are usually extremely patchy in distribution, and therefore also very difficult to sample adequately for a realistic estimate of abundance (Wetzel and Likens 2000). These logistical problems have impeded development of protective numeric nutrient criteria for small to mid-sized streams, which are usually dominated by attached rather than floating algae and plants (Biggs 2000). Although recent advances have been made in sampling filamentous and other, larger attached algae (Stancheva and Sheath 2016), quantitative sampling approaches remain a challenge. Rooted plants are seldom sampled quantitatively because of the extremely labor-intensive effort required to obtain accurate ash-free dry weight data for belowground (roots, rhizomes) as well as aboveground (leaves, stems) structures (Craft 2013).

* Algae (singular, alga; adjective, algal) are primitive plant-like organisms that generally use the same type of photosynthesis as land plants. Algae can be unicellular, filamentous, or colonial microscopic forms (microalgae), or they can be visible to the human eye (macroscopic or macroalgae) consisting of a primitive plant body (thallus) that lacks vascular tissue in most species (thus, they have no roots, stems, flowers, or leaves). Algae are the

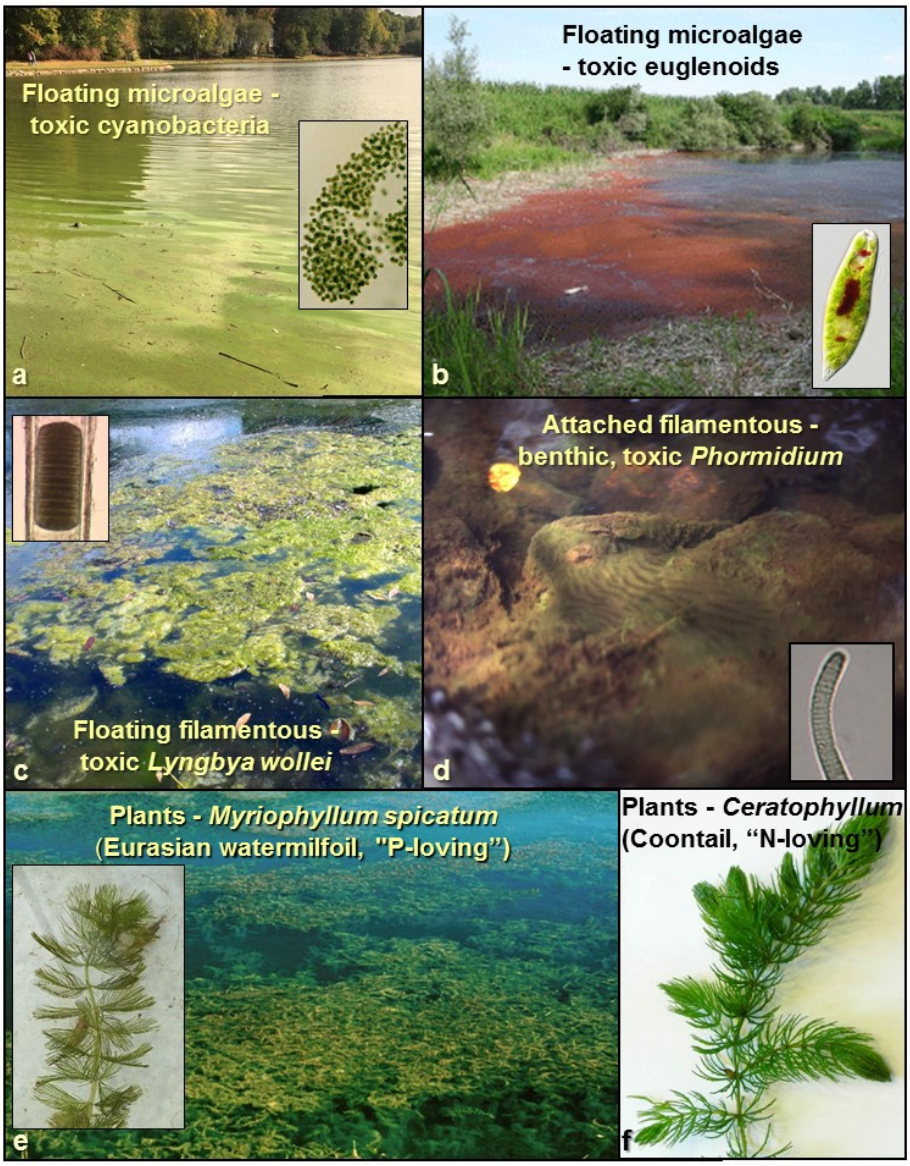


Figure 4. Representatives of the four groups of primary producers known to be most responsive to nutrient pollution: a) Floating microalgae (phytoplankton) - toxic cyanobacteria (inset: Microcystis aeruginosa under light microscopy); b) Floating microalgae - toxic euglenoids (inset: Euglena sanguinea under light microscopy); c) Floating filamentous algae - the toxic cyanobacterium Lyngbya wollei; d) Attached filamentous algae - the toxic cyanobacterium Phormidium sp.; e) Plants - Eurasian watermilfoil (thrives in P pollution); and f) Plants - coontail (thrives in N pollution, especially nitrate).

"grass" of aquatic ecosystems, that is, they form most of the all-important base of the entire aquatic food web. The term has also been used more broadly to include microbes that are not photosynthetic but are closely related to photosynthetic species. And, the term commonly includes cyanobacteria (blue-green algae) because, although they are formally classified as

bacteria rather than algae, they have the same kind of photosynthesis as land plants, and they act ecologically like photosynthetic algae in many ways. Microalgae suspended in the water column, or seston, are called phytoplankton if they have enough time to reproduce before they are carried downstream. Their abundance is commonly measured as the green pigment, chlorophyll a (see primary producers above).

* Floating microalgae (seston, phytoplankton, or suspended microalgae) - This group (see examples in Figure 4a,b) is the best studied of primary producers in aquatic ecosystems (because they are the easiest to sample; see Wetzel and Likens 2000, Wetzel 2001). They discolor the water gold, green, bluish green, brown, etc.; cyanobacteria, for example, often make the water resemble green paint. Suspended microalgae are also the major group of primary producers that is most often considered in assessing whether a waterbody is nutrient over-enriched - even if it is not the most abundant group of primary producers that is responding to the nutrient contamination. That is, growth of primary producers in response to nutrient pollution is commonly considered only from the standpoint of whether there are abundant microalgae suspended in the water column. A serious limitation resulting from reliance only upon measurements of suspended microalgal biomass to assess whether a given water body is degraded from P pollution is that noxious outbreaks of suspended microalgae are notoriously sporadic. They respond quickly to sudden "windows of opportunity" as supporting environmental conditions become available to allow rapid growth in the P-enriched waters (Burkholder 2009 and references therein). Moreover, their outbreaks may last only a few days to a few weeks until a storm, runoff, and increased flushing (decreased water residence time) cause dispersal and washout (Wetzel 2001, Burkholder 2009). Considering that sampling efforts typically are conducted monthly or less frequently, noxious algal outbreaks or "blooms" in response to nutrient pollution can easily be missed (Glibert and Burkholder 2018).

The biomass of the suspended microalgae is usually estimated by measuring an indicator green pigment, the concentration of chlorophyll a (chla), per unit river or lake volume (e.g., chlorophyll a per liter, chla/L). Chla is the "universal plant pigment," found in all plants and plant-like organisms that conduct photosynthesis (Hendry et al. 1987). When samples are taken to measure the water-column chla concentration, formal methodologies generally caution that "particulate debris" such as plant leaves, and even surface scums formed by some of the microscopic algae (Chorus and Bartram 1999), must be avoided. Therefore, the water-column chla concentration, by definition, includes only the suspended microalgae. This fact is important: Chla data completely exclude the biomass of all of the other primary producers in the water body that are responding to the P contamination. This omission of all but the water-column microalgae becomes especially misleading in cases where the water is turbid from high amounts of suspended solids or dark coloration. In such conditions, other groups of primary producers are often the major responders to nutrient pollution because the suspended sediments tend to attract and adsorb microalgae in the water column, leading to their co-flocculation with the heavier particles and settlement down to the bottom (Burkholder 1992 and references therein).

* Floating macroalgae - Floating macroscopic algae (macroalgae, appearing as slimy growth), mostly filamentous forms that are easily visible (Figure 5), are common responders to nutrient pollution in waters which are periodically light-limited from water-column

turbidity (e.g. fine-particulate suspended sediments). Filamentous algae can form raft-like growths that allow them to remain afloat at or near the surface where they can access high light while remaining bathed in the nutrient-rich water. Green algae (Chlorophyta) such as Oedogonium spp. and Spirogyra spp. often form massive, slimy growths at or near the water surface under nutrient over-enriched conditions (Perrin et al. 1988, Benke and Cushing 2005, Mackay 2006, Burkholder 2009 and references therein, Lapointe et al. 2018). Excessive growth of filamentous algae can cause major diel DO fluctuations (Pitcairn and Hawkes 1973, Kirk 1994).

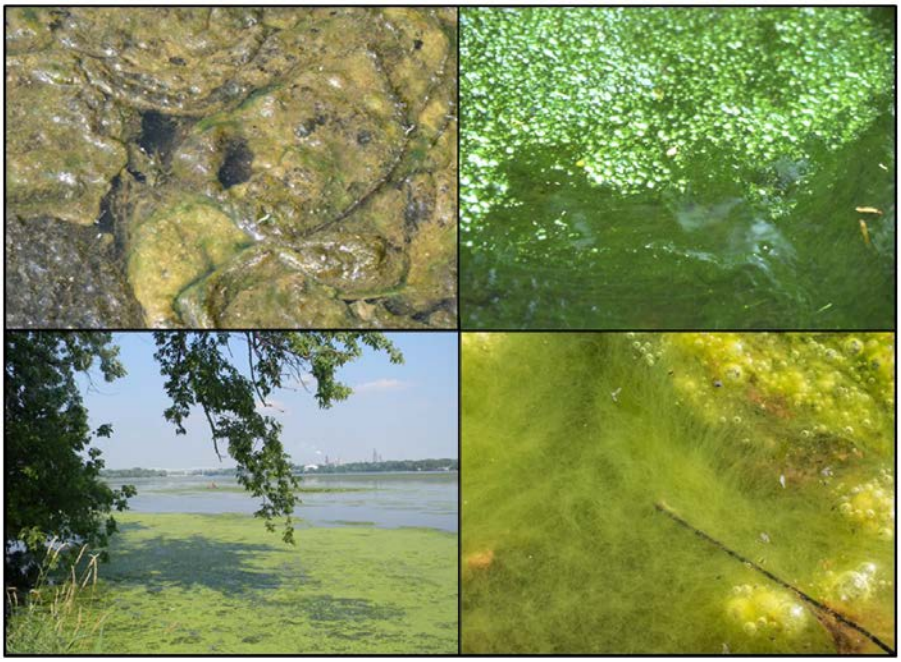


Figure 5. Floating filamentous algae - Upper left: the potentially toxic cyanobacterium Lyngbya wollei; upper right: the green alga Oedogonium; lower left: the green alga Mougeotia mixed with other green algae and duckweeds; and lower right, unidentified filamentous green algae (Groton Lakes Association, Groton, MA – at http://grotonlakes.com/lake-weeds/algae/).

* Attached (benthic) filamentous algae - “Benthic” refers to algae that are anchored (attached) to rocks and other bottom substrata. Attached filamentous algae have an advantage over other benthic algae because they grow with some of their cells extending up into the water column, where light is more accessible (Welch et al. 1988, Graham et al. 2010). Nutrient pollution commonly supports robust growth of filamentous attached algae in shallow waters where light is sufficient to allow them to respond to the elevated nutrient supplies. These algae include some of the most well-known, classic responders to P pollution. In the 1960s, the attached filamentous green alga, Cladophora (Chlorophyta), in Great Lakes Erie and Michigan began to form massive overgrowths along the bottom rock substrata of the western basin of Lake Erie because it was strongly stimulated by P pollution from notable sources such as the Detroit wastewater treatment plant (WWTP) (Lapointe et al. 2018 and references therein, Lake Erie Improvement Association 2012). The Detroit WWTP was the largest source of P entering the west basin of Lake Erie at the time, and agricultural inputs (row crops and industrialized animal production) from the Maumee River also were (and remain) substantial (Bukhari et al.

2015 and references therein). Significant P reduction led to a major decline in Cladophora in the 1970s, until food web disruptions from exotic/invasive zebra and quagga mussels provided an alternate P source that enabled its comeback more than a decade later (note that some external sources have also increased, such as agricultural inputs to the west basin of Lake Erie; Bukhari et al. 2015, Apostel 2016, Ho and Michalak 2017). Cladophora also proliferates on rocky substrata in alkaline streams of the Upper Midwest. In waters of central Illinois, Royer et al. (2008) noted that "it is often filamentous macroalgae…that reach nuisance levels in streams." In the Illinois River basin, a species of attached, filamentous Cladophora was found to grow optimally under high P enrichment (TP concentration 600 µg/L; Leland and Porter 2000).

It is important to note that other forms of benthic algae (sometimes called periphyton) such as microscopic diatoms, green algae, and cyanobacteria, can also form noxious growth as slimy layers (biofilms) in response to nutrient pollution in clear to moderately turbid or murky waters (e.g. Welch et al. 1988, Miltner 2010) until light limits their growth (Wetzel 2001). During warmer seasons in light-sufficient conditions, cold-optimal diatoms generally give way to abundant filamentous green algae and cyanobacteria (Wetzel 2001, Graham et al. 2010, Lapointe et al. 2018).

Harmful algae (all three above groups) - As explained, algae are the major primary producers in most aquatic ecosystems, and they are vital to healthy food webs. Nutrient pollution, however, acts analogously to too much lawn fertilizer that causes overgrowth of plants including noxious weeds. Harmful algae are defined here as algae that are undesirable to people or detrimental to aquatic ecosystems because they (1) become too abundant or "weedy" and "bloom" (proliferate) in response to nutrient over-enrichment and then, at night, use most or all of the oxygen in the water for their respiration, so that the fish and other desirable organisms suffocate or become badly physiologically stressed; (2) become too abundant in response to nutrient enrichment, and overgrow beds of desirable rooted vegetation so that the beneficial plants cannot receive enough light to survive; (3) cause or promote disease in beneficial plants or animals; and/or (4) produce chemical poisons called toxins that hurt or kill fish and other higher animals in the food web, sometimes including humans. Harmful algae of inland waters are highlighted in Burkholder (2009 - general information), Burkholder et al. (2010 - drinking source waters), Lapointe et al. (2018 - cyanobacteria as macroalgae), as general references.

The most notorious and damaging harmful algae in freshwaters, both ecologically and economically, are certain cyanobacteria, which generally form noxious blooms in warmer seasons. Importantly, they are "nutrient loving" - that is, their outbreaks develop in waters that are substantially degraded by nutrient pollution. These organisms are capable of making toxins (cyanotoxins) that have caused serious human disease and death throughout the world. They are well known throughout the Mississippi River basin; in fact, some of the highest cyanotoxin concentrations known were reported from the upper Midwest (Hedman et al. 2008). Cyanobacteria make a wide array of toxins that are capable of making a wide array of other toxins that cause human gastrointestinal illnesses, neurological diseases, skin diseases, and other illness (Burkholder et al. 2009 and references therein). Some cyanotoxins also can promote malignant tumors in mammals (Burkholder 2002, and references therein). The most common cyanotoxins, microcystins, are extremely potent liver toxins (hepatotoxins), and

potent tumor promoters (Burkholder 2002, and references therein). They can cause diarrhea, vomiting, weakness, and death from liver hemorrhaging. They can also damage kidney and lung tissues, as well as nervous system damage; and they can cause increased gene damage (breakage of chromosomes) in human white blood cells. The World Health Organization (2003, p.105) reported "high probability of adverse health effects [from cyanobacterial scum formation] in areas where whole-body contact and/or risk of ingestion/aspiration occur." The WHO listed adverse health effects including the "potential for acute poisoning, potential for long-term illness with some cyanobacterial species, and short-term adverse health outcomes, e.g. skin irritations [and] gastrointestinal illness."

Although the WHO (2003) developed guidelines that related chla concentrations or cell counts to probable cyanotoxin levels, these guidelines have not been broadly applied (U.S. EPA 2016b). The problem is that there is generally no reliable way to determine whether a cyanobacteria bloom is benign, minimally toxic, or highly toxic based on the size of the bloom, the total biomass (as chla), or the number of cells present. Instead, the toxins must be measured. Small blooms and developing blooms tend to be much more toxic than large, aging blooms (Gobler et al. 2016 and references therein). An exception to this generalization is that in some locations, based on the benefit of substantial data, realistic risk assessment of cyanotoxin exposure can be attained. For example, for Minnesota lakes, severe cyanobacteria blooms have been defined as characterized by chla concentrations of 30 µg/L or more (Lindon and Heiskary 2009, p.245). As blooms exceed 30 µg chla/L, the frequency of moderate to high risk of cyanotoxin exposure increases to 12%; above 50 µg chla/L (very severe blooms), the likelihood of moderate to high risk of cyanotoxin exposure increases to 28%; and all high-risk cyanotoxin concentrations have been associated with chla concentrations above 30 µg/L (Lindon and Heiskary 2008, 2009).

* Plants (macrophytes, higher plants, vascular plants) - Plants in aquatic habitats are similar to those on land in having vascular tissue which helps to form their roots, stems, flowers, and leaves. Although they tend to be a minor component of aquatic ecosystems in comparison to algae, they can be abundant locally, and certain species can be major responders to nutrient pollution (e.g., the "phosphorus-loving" invasive species Eurasian watermilfoil, Myriophyllum spicatum; Smith and Barko 1990). These are the most notorious "aquatic weeds," and their rapid overgrowth can choke lakes, block boat traffic, and impair designated uses for primary recreation such as swimming (Figure 6). Their biomass is reported per unit lake or river area, such as grams of biomass per square meter of the bottom mud. Some noxious species, such as hydrilla (Hydrilla verticillata) and Eurasian watermilfoil, can become separated from the bottom substrata to form floating masses in the water column. Then, their biomass is much more difficult to measure and could be reported either on an area (per square meter) or volume (per cubic meter) basis.

III.b.2. Nutrient -Related Processes, Descriptors, and Effects

Trophic status refers to a ranking system for aquatic ecosystems, based on the energy from biota production that is available to the food web (Dodds 2007), or the available amount of organic production (especially phytoplankton production) and nutrient (N, P) levels (Table 1). Trophic state generally refers to the energy available to the food web in an ecosystem There are three basic categories, qualitatively:

Figure 6. Noxious Eurasian watermilfoil clogging lakes, boat motor propellers, and interfering with primary recreation (lower right photo from a Wisconsin lake also includes filamentous algae). Upper left: Lake Taylor, VA (The Virginia Pilot, Aug. 6, 2012); upper right, Oneida County Aquatic Invasive Species Program, Rhinelander, WI); lower left: Candlewood Lake, CT (CT Post, Bridgeport, Aug. 9, 2016; lower right, in a Wisconsin lake (http://shadowlakeassociation.org/wp-content/uploads/2015/10/Milfoil-swimming-through-it.jpg).

Oligotrophic - characterized by relatively high phytoplankton species diversity but low phytoplankton biomass (production), low nutrient supplies (concentrations and loads), low decomposition, and plentiful dissolved oxygen (DO) throughout the water column.

Mesotrophic - characterized by moderate phytoplankton biomass and moderate nutrient supplies.

Table 1. Some eutrophication survey guidelines for lakes, considering average concentrations. From Wetzel (2001).

	Oligotrophic	Mesotrophic	Eutrophic
TP (µg/L)	< 10	10-20	> 20
TN (µg/L)	< 200	200-500	> 500
Secchi depth (m)	> 3.7	3.7-2.0	< 2.0
Hypolimnetic dissolved oxygen (% DO saturation)	> 80	10-80	< 10
Phytoplankton (suspended microalgal) chla (µg/L)	< 4	4-10	> 10
Phytoplankton production (g C m^{-2} d^{-1})	7-25	75-250	350-700

Eutrophic - characterized by high phytoplankton biomass but low species diversity, high nutrient supplies (especially N and P), high decomposition in the bottom water, and bottom-water DO deficits, sometimes also with fish kills.

Hypereutrophic - this fourth category of trophic status is sometimes included, defined as waters with extremely high nutrient concentrations and productivity; Wetzel (2001) describes lakes in this category as having $\geq$ 750-1200 µg TP/L, $\geq$ 100-150 µg chla/L, and Secchi depth transparency $\leq$ 0.4-0.5 m if the turbidity is mostly algal-related rather than from mineral precipitates (e.g., calcium carbonate) or suspended sediments (note - information on this category was not included for TN).

Quantitative rankings for these categories also have been developed, such as Carlson's (1977) trophic state index. This index is calculated using paired data from the upper water column (photic zone, or depth < 2.5 m) for Secchi depth, TP, and chla over the summer season (Carlson and Simpson 1996); resulting index numbers are relegated to general categories of oligotrophic, mesotrophic, eutrophic, and hypereutrophic. The TN concentration can be used to develop a trophic state index as well (Kratzer and Brezonik 1981), although that index was designed to be used in N-limiting conditions which are not characteristic of most surface waters in the Mississippi River basin (Houser and Richardson 2010, Houser et al. 2010).

While rivers and streams are also described using the same terms for trophic status, they are more complex and more variable in response than lakes because they are strongly controlled by (highly variable) flow dynamics. Moreover, lakes are innately much more sensitive to nutrient pollution than streams because water remains in lakes much longer, so that algae and plants have more time to respond to the increased nutrient supplies. Streams and rivers can flush or self-cleanse much more quickly. Thus, much lower nutrient concentrations in lakes, relative to streams, can stimulate noxious algal growth (Dodds and Welch 2000, Dodds 2006) (Table 2).

It should be noted that reservoirs (sometimes called artificial lakes or, in some cases, run-of-river impoundments), made by damming a river, are also complex because they are intermediate between lakes and streams in flushing (water retention time) and in sensitivity to nutrient

Table 2. Suggested guidelines (thresholds) for trophic boundaries in rivers and streams, considering benthic algal abundance (mean and maximum chlorophyll per unit area of stream bottom), suspended (sestonic) microalgal abundance, and mean nutrient (TN, TP) concentrations (data from the 1990s; Dodds et al. 1998). These guidelines were based on cumulative frequency distributions of the data (oligotrophic or nutrient-poor streams, lowest 1/3rd of data; mesotrophic, middle 1/3rd of data; eutrophic, highest 1/3rd of data). An unknown proportion of the sites used to create these guidelines were true reference sites. Note that the numbers for suspended microalgal chla were not specified as to whether they were means or maxima; the those numbers were based on the dataset from Van Nieuwenhuyse and Jones (1996), who reported summer means. Modified from Dodds et al. (1998; also see Dodds and Smith 2016).

Variable	Oligotrophic	Mesotrophic	Eutrophic
TP	< 25 µg/L	25 - 75 µg/L	> 75 µg/L
TN	< 700 µg/L	700 - 1500 µg/L	> 1500 µg/L
Suspended microalgal chlorophyll (chla)	< 10 µg/L	10 - 30 µg/L	> 30 µg/L
Mean benthic algal chla	< 20 mg/m^2	20 - 70 mg/m^2	> 70 mg/m^2
Maximum benthic algal chla	< 60 mg/m^2	60 - 200 mg/m^2	> 200 mg/m^2

pollution. Reservoirs present an additional challenge because they were artificially created after European settlement of this country, so they do not have “natural” background reference conditions per se (Jones et al. 2009). The U.S. EPA (2000a) considers them together with natural lakes as a conservative (protective) approach.

In a waterbody or set of similar waterbodies, ideally the predominant key responder group(s) among those described above should be the main response variable(s) emphasized in the development of numeric nutrient criteria. If `this is not done, then the most important response variable(s) will be omitted from consideration, and the resulting criteria will not protect the designated uses of the system(s) from impairment due to nutrient pollution. For example, if a river downstream from numerous point source dischargers is choked with macrophyte aquatic weeds, but suspended microalgae are used as the key responders to nutrient pollution in the system, then the resulting numeric criteria will fail to restore or protect the designated uses from additional impairment. Such action would be analogous to trying to treat a heart attack by focusing on the stomach; it is not logical, and it would not work.

Overwhelmingly, however, suspended microalgae are usually targeted as the key responders to nutrient pollution in lakes, reservoirs, and larger rivers, and benthic algae (periphyton) are generally targeted as the main response variable in smaller (wadeable) streams (Wetzel 2001, and references therein; Robertson et al. 2006 , 2008; Miltner 2010) - even when other key responders such as floating macroalgae or rooted macrophytes predominate. In part this happens because suspended microalgae are easiest to quantify and the most information is available about them. Less information is available for benthic algae. As examples, recreation and aesthetics designated uses of small, shallow streams and shallow larger rivers in the mountainous western U.S. and British Columbia, Canada were considered to be protected when benthic algal abundance, as the indicator pigment chlorophyll a, was below 50 mg/m^2, and undesirable changes in aquatic life were avoided at benthic chlorophyll below 100 mg/m^2 (Nordin 1985, in U.S. EPA 2000c). Welch et al. (1988) showed that noxious filamentous algae tended to dominate benthic algae in streams when chlorophyll levels exceeded 100 mg/m^2. They described noxious levels of benthic algae in response to nutrient pollution as occurring at 100-150 mg chlorophyll/m^2 of stream bottom. Various other analyses have indicated that benthic algal abundance greater than 150 mg chlorophyll a/m^2 in nutrient-contaminated rivers is undesirable (Heiskary et al. 2013; Suplee et al. 2009, 2015).

Dissolved oxygen (DO) is needed for the survival of most aquatic life, just as oxygen is needed for most organisms on land. It is also needed for most decomposition of dead algal/plant/ animal remains. Oxygen solubility in water is inversely related to temperature (Stumm and Morgan 1995); thus, warmer waters hold less DO, so waterbodies are especially sensitive to low-oxygen conditions during summer. The following conditions can be caused indirectly by nutrient pollution through over-stimulation of algae and algal overgrowth (see harmful algae, above):

Anoxia (adjective, anoxic) - Condition in which there is negligible or no DO in the water (technically 0 mg/L; sometimes considered as less than 0.1 mg DO/L).

Hypoxia (adjective, hypoxic) - Condition in which the water has depressed DO levels that generally are too low to sustain healthy populations of most fish species for extended periods (usually considered as less than 2-3 (4) mg DO/L; Mississippi River Basin

Watershed Nutrient Task Force 2010). Hypoxic levels of 3-4 mg DO/L can stress or kill sensitive stages of some aquatic animals (e.g., Parker 2013).

Diel DO variation (DO "swing" over an entire day-night cycle) - The magnitude of the change in DO over a 24-hour period, from the lowest concentration (often just before dawn) to the highest concentration (often in the early to mid-afternoon during/after high algal/ plant photosynthesis - Figure 7). For example, diel DO swings (variation) of more than 4 mg/L have been shown to be especially detrimental to beneficial aquatic life in rivers of the upper Midwest (Heiskary and Markus 2003).

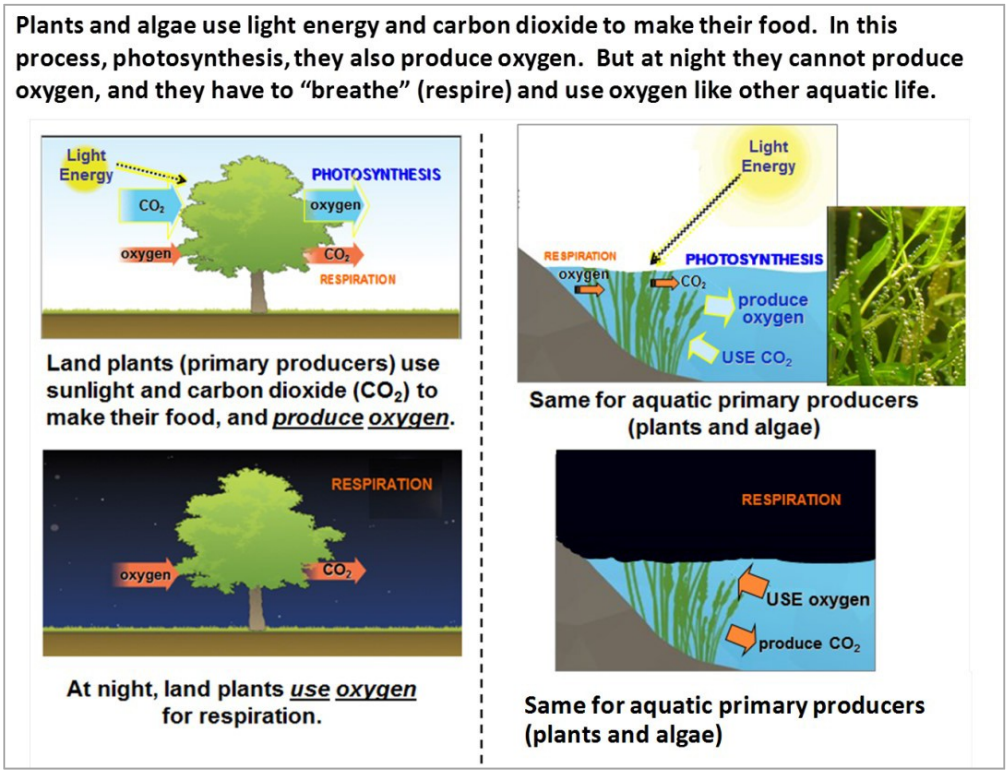


Figure 7. Diagram showing how plants on land (left) and in water (right) affect oxygen through their photosynthesis and respiration during day and night conditions.

Saturation - Condition in which the amount of oxygen dissolved in the water is at equilibrium (that is, the same as) the amount of oxygen in the overlying air. More specifically, oxygen saturation is 100% of air saturation, as the ratio of the concentration of DO in the water to the maximum amount of oxygen that can dissolve in the water at that temperature and pressure under stable equilibrium.

Supersaturation - Condition in which the DO exceeds 100% air saturation, which occurs in surface waters either from rapid flow/bubbling in streams, rapid aeration from overflow at a dam, artificial aeration, or - most commonly - due to algal/plant photosynthesis during the day (Figure 7). Gas supersaturation can cause gas-bubble disease in fish, wherein gas emboli develop in the blood vessels of gill filaments, skin, heart, and other organs, often leading to death (Weitkamp and Katz 1980, Nebeker et al. 1981, Huang et

al. 2010, Suplee et al. 2015 and references therein). Total dissolved gas supersaturation of more than 110-120 (125)% has been recommended as a defensible TDG standard to protect fish from gas-bubble disease (Weitkamp and Katz 1980; see Suplee et al. 2015).

The contribution of DO to TDG pressure varies depending on the barometric pressure. In a trout hatchery study, gas-bubble disease occurred in water maintained at 200% DO saturation, corresponding to 120% total dissolved gases (Edsall and Smith 1991). In the Yellowstone River, Montana, gas supersaturation was found to be driven mostly by diel DO change and peaked at 163% of saturation in model simulations at high nutrient additions, corresponding to 113% TDG saturation as determined with an elevation-based barometric pressure of 690 mm of mercury, and assuming 100% saturation of atmospheric nitrogen and argon gases (Suplee et al. 2015). At the TN and TP concentrations recommended in that study, DO saturation was predicted to be at 144% saturation or less, corresponding to $\leq$ 109% TDG.

Note that one, two, or all three of the above conditions indicating DO stress - anoxia/hypoxia (e.g., DO sag at night), daytime supersaturation, and/or large diel DO variation - can occur in a waterbody affected by nutrient pollution depending on the waterbody characteristics (depth, abundance of key responder group(s), light availability, flow dynamics, etc.), season, and weather events. Many surface waters that are impaired from nutrient pollution show one or two, but not all three, of these conditions in a given 24-hour period (e.g., Huggins and Anderson 2005, Morgan et al. 2006, Tafangenyasha et al. 2010, Wilcock et al. 2010). Diel DO curves often vary greatly within a given season because they are strongly influenced by weather patterns. Moreover, impairment from nutrient pollution commonly happens without any of the above three DO conditions - such as nitrate, nitrite, and ammonia toxicity to aquatic life (Camargo et al. 2005, Camargo and Alonso 2006); increased total organic carbon (TOC) that adversely affects potable water treatment (Walker 1983; and small but highly toxic algal blooms that can develop in response to nutrient over-enrichment when surface waters are oxygen-replete (Boyer 2007).

Nutrient pollution generally does not occur “alone,” that is, in the absence of other pollution. Waters that are impacted by nutrient over-enrichment need protection from nutrient pollution, regardless of whether they are also impacted by other pollutants that can affect DO. As an example, various point and nonpoint sources commonly add biochemical oxygen demand (BOD) as well as nutrients. Waters with both high BOD and high nutrients should be considered nutrient-impaired because numerous studies have shown that, commonly, BOD and nutrients are strongly positively related (Mallin et al. 2006 and references therein). Moreover, some nutrient forms (for example, organic nutrient forms) can directly contribute to BOD or chemical oxygen demand (e.g., ammonia conversion to nitrate in oxygenated waters; Stumm and Morgan 1995).

Nutrient ratio science (nutrient stoichiometry) - The study of changes in the relative proportions of critical nutrients (especially N and P; also sometimes carbon - C, silica - Si) available in the water, in comparison to differences in the allocation of these elements in organisms ranging from phytoplankton to fish and other higher levels of the aquatic food web. Nutrient stoichiometry can also refer to the relative availability of critical elements in the aquatic ecosystem, such as the availability of N and P for algal growth. The water-column

N:P ratio heavily influences the quality of phytoplankton as food for aquatic animals. Nutrient stoichiometry compares nutrient ratios dissolved in the water versus in phytoplankton cells, aquatic plants, and aquatic animals (Glibert et al. 2011 and references therein).

Nutrient ratio science all derives from what is called the Redfield ratio (Redfield 1958), wherein scientists noticed that water-column ratios of several essential nutrients for algal growth were similar to nutrient ratios inside the phytoplankton cells. There are two ways of determining the Redfield ratio, either by atoms or by weight:

Redfield ratio (atomic or molar): 106 Carbon : 16 Nitrogen : 1 Phosphorus

Redfield ratio by mass: 50 C : 7 N : 1 P

Rapidly growing phytoplankton exhibit uptake ratios roughly in accordance with the Redfield atomic ratio, reasonable since the cycles of major and minor nutrients are closely related to biological processes in aquatic systems. The Redfield ratio has been widely used in natural waters as "healthy," that is, in balance with respect to N and P supplies. Significant deviations from it have been interpreted to indicate nutrient stress in systems where one or the other nutrient (N or P) actually was at a supply level that limits phytoplankton growth. For example, prior to European settlement of this nation, water with a N:P ratio of 20:1 (atomic) in low-nutrient, natural waters would have been correctly interpreted to mean that the waterbody was P-limited; the "healthiest" ratio, according to Redfield (1958), would have been a little lower, at 16:1 (atomic). That is, there should be 16 N atoms for every available P atom; in this case, more P would have to be added to bring the ratio down to the "healthy" 16:1. It should also be noted that, while the N:P ratio is often considered as TN:TP, the inorganic N:P ratio (nitrate+ammonium)-to-phosphate ratio is sometimes used (Glibert et al. 2011), so the user needs to clarify which ratio is being applied. In addition, the "healthy" N:P ratio for algae in freshwaters can differ somewhat from the Redfield ratio (Hecky et al. 1993, Geider and La Roche 2002).

Healthy freshwaters are commonly considered to have a N:P ratio at or near Redfield proportions. Rapidly growing algae have N and P uptake ratios that can closely follow the Redfield ratio, expected since N and P cycles are closely related to biological processes. Thus, N and P ratios in the algae can reflect N and P processing in the habitat. Although Redfield ratios are widely applied as "the" only ratio used by healthy algal cells in natural ecosystems, nature is more complicated, so that the optimum ratio for algal growth can vary somewhat depending on the species and the growth rate. At maximum growth rate, the N:P ratio converges to a more narrowly defined range that differs depending on the species (Hillebrand et al. 2013 and references therein). As explained above, however, the supplies (N and P concentrations) as well as the supply ratio are important in assessing the nutrient status of a waterbody. Consider this analogy: A person goes to a restaurant. The restaurant manager apologetically explains that only 1,600 steaks and 100 potatoes are available for the person's dinner. Which will the person run out of first, the steaks or the potatoes? This is obviously a nonsensical question; one person cannot possibly consume 100 potatoes in one dinner. The "Redfield ratio" was followed here (16:1), but both supplies are far too great to be limiting.

Unfortunately, N:P ratios are still being used to infer "which nutrient is limiting," even in extremely nutrient-degraded waters with very high N and P supplies where neither N nor P is limiting. It is done frequently by resource managers considering non-limiting conditions, such as streams heavily affected by partially treated sewage. It is nonsensical to "force" N:P ratios to indicate nutrient-limiting conditions, regardless of what algal bioassays (algal growth potential tests, AGPTs) seem to "indicate," explained as follows.

As an illustration, the TMDL for the Pearl River, Mississippi (Mississippi Department of Environmental Quality [MDEQ] 2009, p.13) misstated, "Point source dominated freshwater systems are generally nitrogen limited." The Pearl River below the malfunctioning Jackson WWTP had high nutrient supplies (1,570 µg TN/L and 390 µg TP/L) and a TN:TP ratio of 4.0, indicating a "sewage signature." According to the U.S. EPA (2000d), minimally impacted conditions or "background" conditions recommended as a basis for numeric nutrient criteria for rivers in this nutrient ecoregion range from 395-618 µg TN/L and 22.5-75 µg TP/L, including subecoregions 65 and 74 within level III ecoregion IX. Based on minimally impacted conditions, both N and P were excessive. Most of the TN was ammonia; some of it was oxidized to nitrate. Both N forms are readily used by algae (Wetzel 2001 and references therein). Most of the TP was phosphate, the preferred form of P used by algae (Wetzel 2001). These inorganic N and P concentrations were much higher than levels that would have limited algal growth. Nevertheless, based on algal growth potential tests (AGPTs; Raschke and Schultz 1987), MDEQ (2009) asserted that the Pearl River was nitrogen-limited. Because sparse historic information had suggested that P should have been limiting rather than N, P reductions from the point sources were targeted while the high N levels were not addressed. Clearly, both TN and TP should have been decreased as a scientifically sound management strategy because the supplies were much higher than "limiting" (oligotrophic to mesotrophic) values for rivers. This lack of understanding, that nutrient limitation cannot be invoked unless nutrient supplies are actually at limiting levels, resulted in a TMDL that will continue to cause nutrient-related degradation of the Pearl River, including both excessive N and extreme N:P stoichiometric imbalance.

MDEQ (2009) based its assessment on nutrient ratios (while overlooking the excessive N and P supply concentrations) and AGPTs, wherein there was a higher algal response with N additions than with P additions in laboratory assays. The mistake made regarding interpretation of the AGPTs rests was a fundamental failure to understand that in an extremely high nutrient regime such as the Pearl River below the Jackson WWTP, the suspended microalgal assemblage would have shifted long ago (years) to highly pollution-tolerant, noxious species like the responders to nutrient pollution described above. All that the AGPTs provided was information on which nutrient would fuel yet more growth of those noxious, undesirable species; the natural assemblage had long ago been driven out by the noxious, highly pollution-tolerant taxa.

The U.S. EPA (2015a) recognized that restoration of waters which have been driven far out of balance in their nutrient stoichiometry and their nutrient supplies will require reducing both P and N pollution (that is, N and P co-management) toward re-establishing natural (reference) N:P ratios. Unfortunately, as illustrated in the above example, resource managers have often continued to emphasize control of one nutrient while minimizing control of the other. It has also been argued that if resources are not available to reduce both N and P pollution to

desirable proportions, then nothing should be done (!). Obviously, it is important to do as much as possible, now, to reduce nutrient pollution; if there is funding to reduce only one nutrient at present, then that should be done while working to obtain the funds to reduce the other.

Direct versus indirect effects on aquatic life - The above key responders to nutrient pollution (Section III.b.1) are directly stimulated by increasing nutrients until nutrients no longer limit their growth. Many impacts of nutrient pollution, however, are indirect rather than direct, and result from modifications to aquatic food webs due shifts in the food web to more tolerant/ less desirable species of primary producers (Burkholder and Glibert 2013, and references therein). Examples of indirect effects include lower food quality at the base of the food web, affecting all food web levels; oxygen extremes, leading to disease and death of beneficial biota; increased turbidity and loss of critical habitat species; and clogged filters and drinking water hazards from high algal/plant biomass and algal toxins (Burkholder and Glibert 2013). The extent and severity of indirect impacts depend on (i) the primary producers initially present; (ii) the characteristics of the enrichment (amount, N:P ratio balance, nutrient form, and the history [recent vs. chronic, continuous vs. pulses]); (iii) the flushing rate or water residence time; and (iv) other environmental factors that affect the primary producers (temperature, light/turbidity, season, grazers, pathogens, other pollutants). For example, if nutrients are added to a river during a period of high precipitation, the dilution/washout can minimize the suspended algal response. On the other hand, nutrients added to the same area of the river during a low-flow period on bright, sunny days can stimulate a noxious algal bloom.

Numeric nutrient criteria are relatively straightforward when an aquatic system shows a clear cause-effect response, for example, when a nutrient addition causes an obvious increase in algal biomass to an undesirable level - but many serious impacts of nutrient pollution are more insidious, indirect effects. Indirect effects especially occur from chronic nutrient pollution delivered as pulses day after day or week after week for months or more, which commonly result in gradual, long-lasting, major impacts on aquatic communities. The challenge facing resource managers is to develop numeric criteria that protect the designated uses of the waterbody(s) from impairment due to indirect as well as direct effects of nutrient pollution.

Effects on recreational uses – The World Health Organization (WHO) (2003), various states (Stephenson 2007), and the U.S. EPA (e.g., 2016b) have also considered the impacts of nutrient pollution on recreational uses, as illustrated by the following examples. The underlying rationale for these efforts is that nutrient pollution strongly stimulates growth of noxious cyanobacteria (blue-green algae), which generally become the dominant or most problematic primary producers in many surface waters nutrient pollution increases. In efforts to protect human health from toxic cyanobacteria, the most important information needed is the amount/type(s) of toxin present (U.S. EPA 2016b). Unfortunately, however, health safety managers across the U.S. have struggled to obtain quantitative information about the presence of cyanotoxins in recreational waters because several major groups of these chemicals can be produced by the same bloom, at the same time or at different times; toxin concentrations are often higher at the beginning of a bloom when cyanobacteria are less abundant than later in the bloom (Gobler et al. 2016 and references therein; but see Lindon and Heiskary 2008, 2009 below); and accurate concentrations generally are much more expensive to measure than simple metrics for cyanobacteria abundance, such as cell number or biomass (chla concentrations).

Scientific understanding about toxin presence and impacts during cyanobacteria blooms is still remarkably limited, considering that little is known about impacts on human health from aerosol exposures or from interactive effects of co-occurring cyanotoxins (U.S. EPA 2016b).

Cyanobacteria abundance and/or a common cyanotoxin – The WHO (2003) suggested cyanobacteria-related thresholds for low, moderate, high, and very high probability of human health effects. The cyanobacteria metrics that were considered included cyanobacteria abundance as cell numbers per milliliter, cyanobacteria abundance as biomass indicated by chla concentrations, and concentrations of a common microcystin (MC), the cyanotoxin MC-LR. Cyanobacteria abundance below 20,000 cells/mL and/or 10 µg chla/L (biomass indicator), and/or MC-LR concentrations below 10 µg/L, corresponded to Low probability for human health effects. Expectation of Moderate occurrence of health effects coincided with cyanobacteria abundance at 20,000 to 100,000 cells/mL, 10 to 50 µg chla/L, and/or MC-LR at 10 to 20 µg/L. High probability for health effects was linked to 100,000 to 10 million cells/mL, 50 to 5,000 µg chla/L, and/or 20 to 2,000 µg MC-LR/L. Very high probability for health effects coincided with > 10 million cells/mL, > 5,000 µg chla/L, and/or > 2,000 µg MC-LR/L.

These WHO thresholds were developed considering that regulatory agencies commonly have the funding or facilities to measure only one of the three factors – usually suspended microalgal chla or, less frequently, an accurate measure of cyanobacteria cell numbers (Wetzel and Likens 2000). Unfortunately, there are major limitations of this approach, most importantly, that small cyanobacteria blooms can be highly toxic (Boyer et al. 2007, U.S. EPA 2016b) – that is, cyanobacteria cell numbers and biomass as chla actually are often poorly related to cyanotoxin concentration. Secondly, as explained above, many regulatory agencies across the nation do not have the laboratory capabilities and/or the funding to measure cyanotoxins such as MC-LR, and cyanotoxins are much more important from a human health perspective than cyanobacteria abundance (U.S. EPA 2016b). A third major limitation of this approach is that cyanobacteria produce many different toxins (e.g., there are more than 100 forms of MCs alone) – and even the same cell can produce multiple toxins (Burkholder et al. 2018 and references therein) – which suggests that measurement of only one common toxin (e.g., MC-LR) or group of toxins (e.g, MCs) that may or may not be present in a given bloom is inadequate to safeguard human health from serious cyanotoxin exposure. A fourth major limitation is that excessive concentrations of chla (up to 5,000 µg/L or more) and MC-LR (up to 2,000 µg/L or more) that the WHO linked to high/very high potential for human health effects generally occur as “scums,” that is, as large clumps of cyanobacteria at the water surface and along shorelines. Yet, most state agencies require samples collected from below the water surface and shorelines. The result is that these highly toxic scums are not included in the samples analyzed, despite the fact that scums are a known major threat to the health safety of recreationists, especially children (Chorus and Bartram 1999, WHO 2003).

In a related approach, but specific to a group of waterbodies, chla concentrations were related to the presence of MC levels to enable use of chla as an indicator for potential risk to human health from MC exposure. This work was conducted in a set of Minnesota lakes. Based on the data, “severe” and “very severe” cyanobacteria blooms linked to MCs were defined as being above mean chla concentrations of 30 µg/L and 50 µg/L, respectively (Lindon and Heiskary 2008, 2009). All MC concentrations that posed high risk to human health were found, in those lakes, at chla concentrations above 30 µg/L.

The U.S. EPA (2016b) released draft recommended quantitative limits for MCs (4 µg/L) and cylindrospermopsin (8 µg/L) in recreational waters, unrelated to chla concentrations (also see U.S. EPA 2015b-d). These suggested limits, based on oral ingestion only, are guidelines rather than numeric standards. As advisories, they are not enforceable.

User perception surveys – Various states have considered the potential utility of recreationists' perceptions about the suitability of waterbodies for swimming, boating, and aesthetics in setting numeric nutrient criteria. These surveys have been based on water clarity (which can be strongly affected by algal blooms), TP concentrations, and/or suspended microalgal chla (e.g., Heiskary and Walker 1988, Smeltzer and Heiskary 1990, Smith et al. 1995, Hoyer et al. 2004). The studies are based on simple surveys that solicit respondents' subjective evaluation of the water for recreational use. Respondents rank the waterbody on a five-point scale that is simply defined, ranging from 1 = "beautiful" or "excellent" to 5 = "not useable" (e.g., Figure 8). Statistical analyses are used to correlate the respondent rankings and the water quality conditions that existed at the place and time when the survey was conducted.

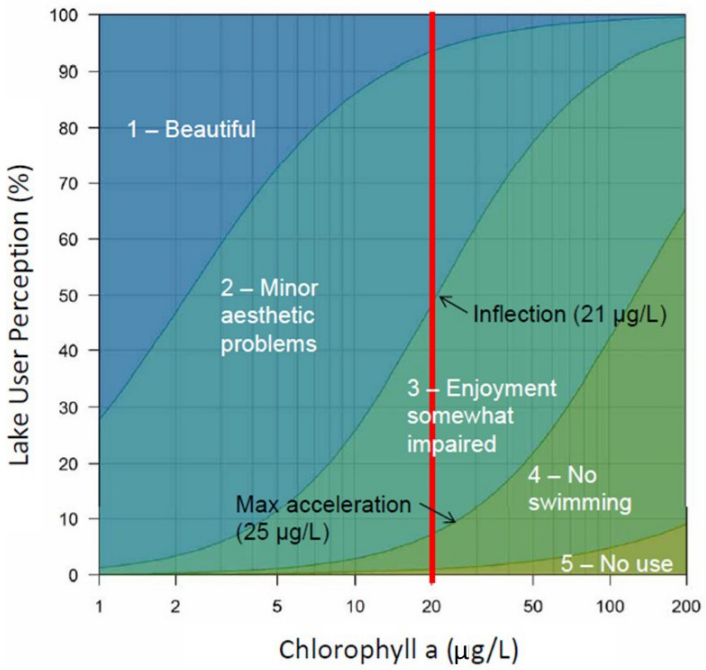


Figure 8. User perceptions of the suitability of Wisconsin lakes for recreation, based on rankings (1-5, shown in white) in surveys. The inflection point at ~20 (21) µg chla/L was used by the Wisconsin Department of Natural Resources to define "nuisance" algal blooms. From Minahan (2017).

As examples, a user perception survey indicated that half of recreational respondents in Wisconsin perceived some impairment to lake use, and some would not swim, at algal levels higher than ~20 µg chla/L (Figure 8). The information was used by the state environmental agency to support a definition of "nuisance" algal blooms at 20 µg chla/L or higher (Minahan 2017). For a set of reservoirs in southeastern Texas, 72% of more than 1500 participants ranked waters at or below ~15 µg chla/L in the top two categories ("beautiful, could not be nicer" or "very minor aesthetic problems") (Glass et al. 2006). In West Virginia, recreationists' evaluations of lake quality were correlated with surface-water TP concentrations; at TP levels below ~25 µg/L), most respondents assessed lakes as excellent to generally nice for recreation (rank, 1-2) (Hansen et al. 2006).

An inherent drawback in the utility of user-perception surveys is that the participants sometimes show a fundamental inability to relate algal biomass (chla levels or "green water") to potentially dangerous conditions for their health. Thus, in some surveys, recreationists have assessed waters with low chla concentrations as moderately poor in recreational quality; in others, such as the Minnesota lakes described above, participants have assessed waters with ~30 µg chla/L as the top ranking ("beautiful, could not be nicer") (e.g., Force and Macbeth 2002, Minnesota Pollution Control Agency 2010; and see Burkholder 2010). Moreover, these studies mostly have not been based on representative samples of recreational users. Instead, most respondents have been state agency staff or members of volunteer monitoring associations. Thus, "the results do not reflect representative cross-sections of recreational users, or the intensity of recreational use across different sites" (Stephenson 2007).

An important generalization from these studies is that the level of water quality assessed by recreationists as suitable for swimming varies substantially with region, waterbody, and the individual users. The evaluation strongly depends on what people are used to seeing; as Hoyer et al. (2004, p.247) wrote, "Lake users located in areas dominated by eutrophic lakes are more likely to tolerate green water and consider it good water quality," and the opposite has also been found (Smeltzer and Heiskary 1990). User perceptions of water quality may also be more sensitive to a change in water quality than to absolute parameter levels (Michael et al. 2000).

III.b.3. Units

The most common units for P and N concentrations are:

milligrams per liter (mg/L, mg L^{-1}; 1 milligram = 1/1000th of a gram) or
micrograms per liter (µg/L, µg L^{-1}; 1 microgram = 1/millionth of a gram, or 1/1,000,000th of a gram).

Scientists also use: ppm, parts per million, wherein 1 ppm = 1 mg/L.
ppb, parts per billion, wherein 1 ppb = 1 µg/L.

For suspended algal (phytoplankton or seston) biomass as chlorophyll a, values are usually reported in µg/L but may instead be reported as milligrams per cubic meter (mg/m^3), wherein 1 µg/L = 1 mg/m^3. Chlorophyll a, representing the living algal biomass, can be corrected using simple laboratory techniques (Wetzel and Likens 2000) to exclude the degraded chlorophyll (pheopigments) in dead algae, which can sometimes cause spuriously high estimates of the living algal abundance.

The N and P loads are also an important consideration in protecting surface waters from excessive nutrients, for example, in efforts to develop total maximum daily loads (TMDLs) that are required under some circumstances by the Clean Water Act or permits for sewage treatment plants (below). The external load (= volume input x concentration in that volume) coming in from the watershed is usually expressed as weight (e.g., grams, pounds, tons, tonnes) per unit watershed surface area (hectare or acre, square mile or square kilometer) on a per-season or per-year basis. The airshed also contributes, sometimes substantially, to the external nutrient load. Internally, that is, within the waterbody, some cyanobacteria species can convert ("fix") atmospheric nitrogen gas (N_2) into ammonia (NH_3), which can be used by many algae and plants; thus, it is important to consider the role of nitrogen-fixing cyanobacteria, which can be substantial (Wetzel 2001 and references therein). The total load must also consider

the N and P contributed internally from the lake or river sediments into the overlying water, referred to as the N and P sediment flux. The bottom sediments are considered the "memory" of an aquatic system because a major proportion of the nutrients added to the waterbody from external sources settle out to the bottom, adsorbed to suspended sediment particles, algal cells, etc. (Wetzel 2001). Nutrients accumulate in the bottom sediments, and when the water column mixes or when there is hypoxia/anoxia in the lower water column, a portion of the N and P is released from the sediments and moves up into the water. This sediment source of N and P is sometimes called "legacy" nutrients. The total N and P load is the combination of the external load + the internal load (within the waterbody).

III.b.4. Statistics and Models

Calibrate (a model) – estimate and adjust model parameters to improve the agreement between model output (predictions) and actual data for the system (Rykiel 1996). A model typically is calibrated by adjusting parameters for model equations within physically defensible ranges until resulting predictions give the best possible fit to the observed data from the system (U.S. EPA 2009a). Complex models that simulate more processes require more actual data from the system being modeled to calibrate reliably (McConkey et al. 2004), that is, to ensure that it has predictive power.

Data – actual measured values or conditions, such as water quality monitoring data.

Linear relationship and correlation – Linear regression analysis emphasizes predicting one variable from the other. Assuming that the fundamental assumption is correct - that the relationship between the data being analyzed is linear - the correlation coefficient indicates the strength of the relationship between the two variables (Sokal and Rolff 2012).

Parameter – a constant (or coefficient) in a mathematical equation that is part of a model. The predictive reliability of a model largely depends on the coefficients used to parameterize it (= the process of selecting the parameters and assigning realistic values to them for the specific system being modeled).

Parameterize – the process of building the model, wherein parameters are selected and realistic values are assigned to them considering the specific system of interest.

Recruitment – the addition of new individuals to a population (e.g., via reproduction and growth), or to successive live history stages within a population (Caley et al. 1996).

Validate (a model) – use different data from the system being modeled to show that the model has a satisfactory range of accuracy for its intended application. Validation compares modeled (= simulated) system output (for example, dissolved oxygen concentrations in a river) with actual data for the system. These data cannot have been used in model development or calibration (Rykiel 1996, Mazzotti and Vinci 2010).

III.c. Decreasing Nutrients to Limiting Levels for Protective Criteria

Surface waters historically contained all of the essential elements needed for algae and plants to have balanced growth, and P, or N and P, were the nutrients that these primary producers ran out of first, relative to their physiological needs (Hecky and Kilham 1988, Wetzel 2001).

Thus, in freshwaters, P (or N + P) were evaluated as the "primary limiting nutrients" or "co-limiting nutrients" for algal and plant growth. Where other conditions needed for photosynthesis are available, such as adequate light and a water residence time that is long enough to allow growth without being removed from the system, algae and plants are directly stimulated by P or N+P to grow.

If N and/or P become available in amounts that are in excess of their physiological needs, eventually algae and plants can become P- or N- or N+P-"saturated." In such situations, another essential resource (such as another nutrient, or a physical resource such as light) can become more important in controlling algal/plant growth. This concept is very important because when a system has been degraded with chronic, high nutrient pollution, that system has to be "pushed back" to P and N levels that are within a limiting range before noxious plant/algal biomass will decrease and the system can begin to recover (Figure 9). As shown, there is no further growth response of the algae to increasing nutrient concentrations at the right of the red arrow; the curve is flat. If a surface waterbody has N or P concentrations within the range indicated in orange, then management actions to reduce the nutrient concentration to a point still within the orange range will not accomplish a reduction in the algal/plant biomass because they will still be N- or P-saturated. The primary producers will only decrease when the nutrient levels are decreased back to levels that limit their growth, here, within the blue box to the left of the red arrow.

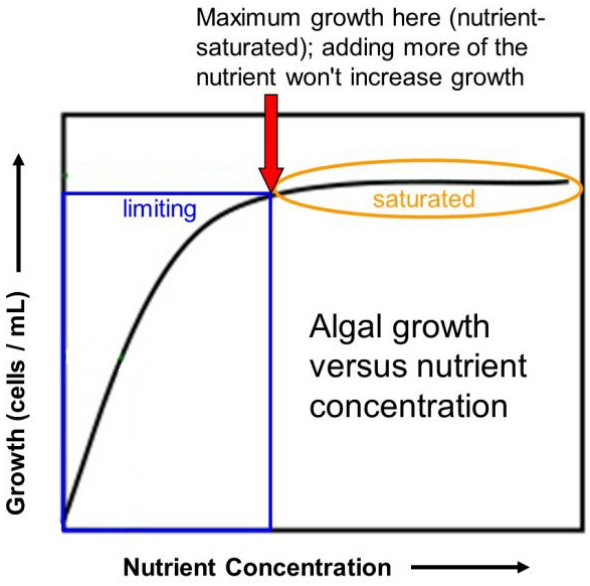


Figure 9. Generalized diagram of algal growth versus nutrient concentration. The left portion of the curve (within the blue box) is where the nutrient is at limiting concentrations, so there is a strong response with growth increasing as the nutrient concentration increases. In contrast, the flat area of the curve (to the right of the red arrow, within the orange circle) is where the nutrient concentration is so high that algal growth is "saturated," that is, there is no further increase in growth as the nutrient concentration continues to increase. Management actions that reduce the nutrient concentration from one point within the orange circle to another point that is still within the orange circle will not noticeably reduce noxious algal blooms. Instead, the nutrient concentration would have to be reduced to an area of the curve within the blue box in order to achieve the desired noticeable decrease in noxious algae.

This concept was further developed by scientists who examined data from nearly 300 temperate streams and rivers worldwide, and found that there was a strong curvilinear relationship between summer average chlorophyll concentration for suspended microalgae and summer mean TP concentration (Van Nieuwenhuyse and Jones 1996 - Figure 10; note that small streams sampled within a few miles of treated sewage outfalls were omitted from consideration because, as illustrated from the above example of the Pearl River, the extremely high inorganic P concentrations and, likely, the high algal biomass that responded, would have biased the study to emphasize point source conditions). Also note that other groups of algal/plant responders

were not considered. The drainage basin area had a strong effect on chla concentrations at all TP levels, as indicated in Figure 10. This study, supported by other work (Royer et al. 2008 - Illinois streams), indicated that the cutoff between P limitation and P saturation for suspended microalgae in rivers and streams occurs within the range of 200-250 µg TP/L. The analyses indicate that at higher TP concentrations, management actions to reduce the average summer TP concentration will not result in noticeable declines in noxious algal blooms unless TP is reduced to a limiting level for algal growth. Although similar analyses are not available for TN or inorganic N, N would be expected to behave analogously; that is, N pollution would need to be reduced back to limiting levels for algal growth in order for managers to achieve a noticeable decline in noxious suspended microalgal blooms.

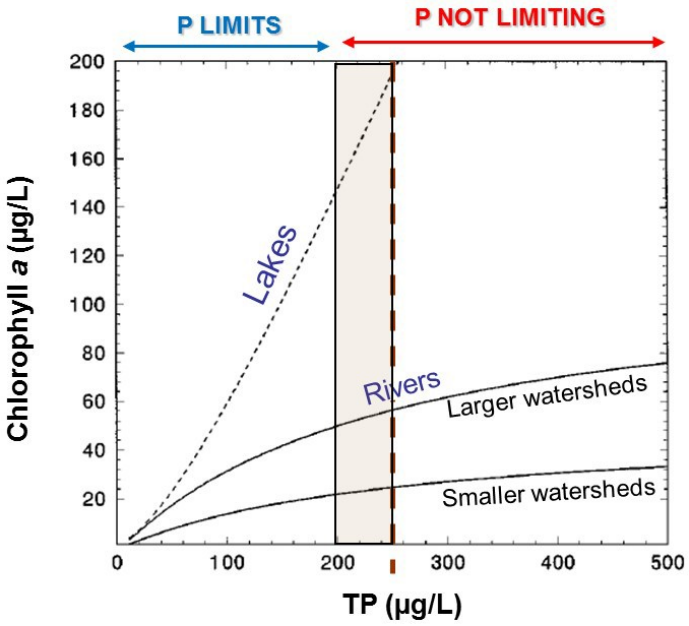


Figure 10. Watershed level, considering TP in streams and rivers, with suspended microalgal biomass represented by the chla concentration (note that other groups of algal/plant responders were not considered). At TP concentrations higher than 250 µg/L (dashed line), the curves for rivers in both larger and smaller watersheds are nearly flat - that is, the P pollution is so high that it no longer limits algal growth and there is minimal algal response. The gray bar represents the range indicates the range at which the algae become P-saturated; in this study by Van Nieuwenhuyse and Jones (1996), the saturation (lack of further limitation by P) occurred at ~250 µg TP/L, whereas in another study conducted in Illinois (Royer et al. 2008), concentrations above 200 µg TP/L were not limiting. Graph modified from Van Nieuwenhuyse and Jones (1996).

In various lakes and rivers throughout the world, reductions in excess nutrient supplies have depressed algal biomass (e.g., Jeppesen et al. 2005, Meals et al. 2010, Rook 2012, Schindler et al. 2016) - usually after a lag period (months to ~a decade) when nutrient-rich bottom mud is slowly buried so that the "legacy" nutrients accumulated there no longer can enrich the overlying water to promote noxious algal blooms. The information shown in Figures 9 and 10 has major ramifications for developing protective numeric nutrient criteria, whether at general (ecoregion) or local (wastewater treatment plant) scales.

IV. Numeric Nutrient Criteria: Rationale, Lines of Reasoning, and Approaches

IV.a. Rationale - The main rationale for setting numeric nutrient criteria is that they create clear, direct goals aimed at the two key parameters (N and P) that, in contaminated waters, repeatedly have caused undesirable impacts. They facilitate improved control of both point and nonpoint sources of nutrient pollution (U.S. EPA 2009b). The U.S. EPA (2000a-c) recommended focus on the causal variables, nutrients TN and TP, and secondarily on response variables such as chla, in developing numeric nutrient criteria.

It is important to emphasize that the overall objective is to develop numeric nutrient criteria which protect the designated uses of the waterbody(s) (e.g., primary or secondary recreation, potable water supply, aquatic life) from impairment due to nutrient contamination. The criteria need to be protective; they should enable protection of the waterbody(s) before noxious algal blooms, DO sags, large diel DO swings, kills of beneficial aquatic life, and other eutrophication-related impacts occur, not afterward (Stoner 2011, U.S. EPA 2013).

IV.b. Two Lines of Reasoning - The U.S. EPA (2000a-c) recommended developing numeric nutrient criteria following one of two lines of reasoning. Under the "minimally impacted" line of reasoning, the user estimates the natural or least altered levels of nutrient stressors and response variables in nutrient ecoregions, that is, areas characterized by similar soils and other background conditions. Criteria are based on the premise that nutrient levels should be similar to the natural or least altered conditions.

In contrast, the "cause-and-effects" line of reasoning focuses on designated uses of the waterbody(s) of interest, wherein stressor-response field data, experiments, and/or modeling are used to assess or predict what the numeric nutrient criteria should be in order to protect the most sensitive designated use(s). Within each line of reasoning, various data can be used following science-based approaches (Section IV) or misused (i.e., not used as intended - Section VI), as described below.

The cause-and-effects line of reasoning may identify higher nutrient concentrations than reference conditions that still protect designated uses; indeed, the U.S. EPA (2000a-c) sanctioned use of minimally impaired conditions which, by definition, would have higher nutrient concentrations than reference conditions. For example,

> We often expect algal species richness and evenness to decrease in streams with increasing pollution, but nutrients also are a resource stressor. Increasing resource availability from low to medium levels can increase diversity because low nutrient concentrations might constrain algal taxonomic composition to those species adapted to low nutrient concentrations….Thus, release from nutrient limitation might allow increases in species richness in physically disturbed systems such as streams (Biggs and Smith 2002, Huston 2004). Despite this potential increase in overall species richness, addition of nutrients could reduce relative abundance and richness of native species by altering competitive hierarchies or habitat structure (Huston 1994). - Stevenson et al. (2008)

Caution is warranted, however, regarding adoption of criteria that allow pollution significantly above natural or reference levels. For example, protection of a generally described "aquatic life use" designation may not be protective of species that are sensitive to nutrient pollution.

Regardless of the approach selected, care should be taken to emphasize populations of sensitive species known to be endemic to the waterbody(s) and intolerant of nutrient pollution impacts, rather than targeting protection of pollution-tolerant species that can survive in waters affected by moderate to high nutrient contamination (e.g., Stevenson et al. 2008, Robertson et al. 2008). The selected approach should also use high-quality data (see below) from the past ~decade, produced by state-certified or other specialist laboratories and collected at least

monthly. As a third point of counsel, many waterbodies have multiple designated uses, and numeric nutrient criteria should be developed for the designated use that is most sensitive to adverse impacts from nutrient pollution (U.S. EPA 2000c, 40 CFR §131.11(a); U.S. EPA 2013).

IV.c. Five Approaches - Five approaches have been recommended by the U.S. EPA (2000a,b) to develop numeric nutrient criteria. These approaches explicitly or implicitly use one or the other or both of the two lines reasoning discussed above.

Minimally impacted - The reference approach falls under the "minimally impacted" line of reasoning, wherein criteria are based on the water quality conditions of least or minimally disturbed waters. The responses of many beneficial biota to changes in nutrient concentrations are non-linear, such that a small change in nutrient concentration corresponds to a relatively large change in the biota. The final nutrient criteria should be selected so as to minimize degradation in the biotic integrity of the waterbody(s) (Robertson et al. 2006). The threshold nutrient concentration (below) where this large, often undesirable change occurs (e.g., loss of sensitive beneficial biota) is often only slightly higher than the reference concentration (e.g., Robertson et al. 2008). Thus, the reference approach, used protectively, favors anti-degradation.

Causes and effects - One or more of the other four approaches can be used within the "cause-and-effects" line of reasoning:

The literature approach relies on suggested boundaries for enrichment classification of lakes and reservoirs, and of rivers and streams, developed by scientists from observations of waterbodies in a given geographic region or climatic zone.

The stressor/response-based modeling approach uses available field and experimental data to estimate statistical relationships between nutrient concentrations and response measures that are relevant to the designated uses.

The mechanistic modeling approach develops mathematical representations of the ecological system(s), processes, and parameters (variables) using equations that can be calibrated using site-specific data. This "ecosystem-level" modeling approach is more complex than stressor-response modeling because it captures various biogeochemical and other processes that characterize the entire ecosystem. Mechanistic models can mathematically describe interactions between nutrient availability, algal nutrient uptake, light, flow, etc. that define relationships between nutrient concentrations and biological and water quality responses at a given location; and they can also simulate water quality variables affected by nutrient pollution, such as dissolved oxygen and pH (Suplee et al. 2015).

Finally, "hybrid" analyses using more than one approach, can include the reference (minimally impacted) condition, "weight of evidence" of different statistical assessments of stressor-response relationships, and mechanistic models (U.S. EPA 2000a-c). These five approaches are described in more detail below.[1]

[1] This writing does not consider the more complex guidance published for developing numeric nutrient criteria for estuaries and coastal waters (U.S. EPA 2001b).

Applicable to all cause-and-effect approaches, the many nutrient stressor/response relationships that have been reported may be broadly considered within two general categories: Some show a clear threshold or abrupt change along the nutrient concentration gradient, whereas others show a gradual decline as the nutrient concentration increases, so that a threshold is not clear. For cases where a threshold TN or TP concentration is evident, then to protect most aquatic life (including a reasonable number – some, albeit not all – pollution-sensitive species), the idea is to "stay away from the cliff." That is, the threshold concentration should not be exceeded. In contrast, if the decline in a desirable biological metric is not abrupt but, rather, more gradual as the TN or TP concentration increases, then a numeric nutrient criterion should be set which protects most of the species. The basic guidance used by the U.S. EPA (Stephan et al. 1985, p.17) for setting other substance criteria should be analogously applied to assess where to "draw the line" –

> Instead of being [set as] equal to the Final Acute Value, the Criterion Maximum Concentration [here, a numeric nutrient criterion] is now [set as] equal to one-half the Final Acute Value. The Criterion Maximum Concentration [here, the numeric nutrient criterion] is intended to protect 95% of a group of diverse genera (unless a commercially or recreationally important species is very sensitive). However, a concentration that would severely harm 50% of the fifth percentile or 50% of a sensitive important species cannot be considered to be protective of that percentile or that species. Dividing the Final Acute Value by 2 is intended to result in a concentration that will not severely adversely affect too many of the organisms.

Application of this rationale to numeric nutrient criteria would result in criteria that protect 95% of the species diversity of beneficial biota such as fish and benthic macroinvertebrates, including some (although not all) sensitive species (see Section IV.d below).

1. Reference condition (minimally impacted, or percentile) - This approach is based on water quality data for TN and TP in freshwaters within U.S. EPA nutrient ecoregions and sub-ecoregions, mainly developed and delineated based on soil types and the predominant land use (Figure 11; and see Section 304(a) of the Clean Water Act, national nutrient criteria recommendations - https://www.epa.gov/nutrient-policy-data/ecoregional-criteria).

Criteria for each of the delineated ecoregions and sub-ecoregions are derived from the range of nutrient levels characteristic of reference or minimally impacted waterbodies, defined as the least disturbed or minimally disturbed conditions within a nutrient ecoregion that support designated uses (U.S. EPA 2010). The most pristine waters (i.e., with lowest P and N levels) within a nutrient ecoregion or sub-ecoregion will occur as a range of concentrations. The U.S. EPA recommended taking the 75th percentile of the range of TN and TP concentrations from the reference or minimally impacted waters (past ~10 years of data) within the sub-ecoregion was recommended for use as the numeric TN and TP criteria. That is, after developing a distribution curve of TN and TP concentrations in a group of reference waters, the criteria are set so that 75% of the observed concentrations are below the numbers chosen for the criteria (Figure 12).

Further, the U.S. EPA reasoned (dubiously – see Evans-White et al. 2014) that the 75th percentile of data from true reference waters should be similar in TP and TN concentrations as

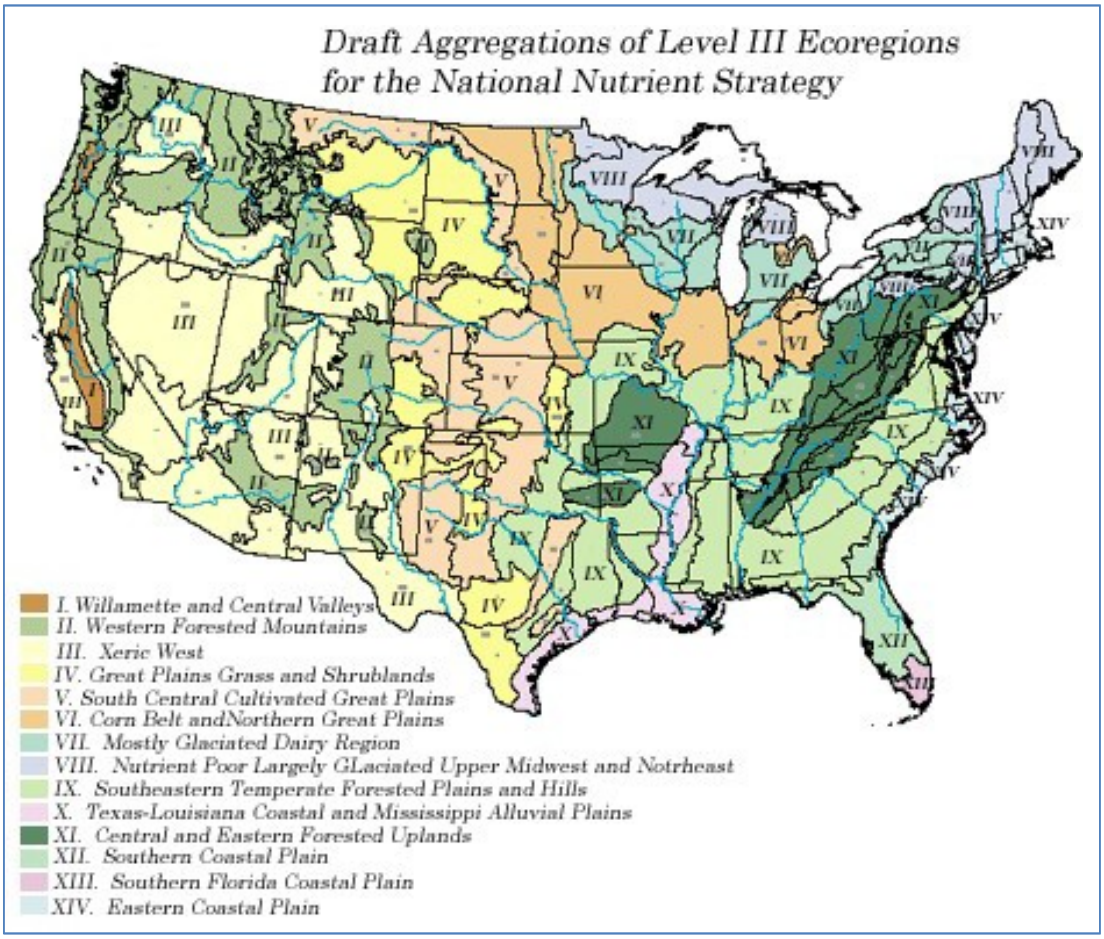


Figure 11. Map showing the level III nutrient ecoregions developed by the U.S. EPA (2000a) for its statistical approach for derivation of recommended numeric nutrient criteria for freshwaters across the nation.

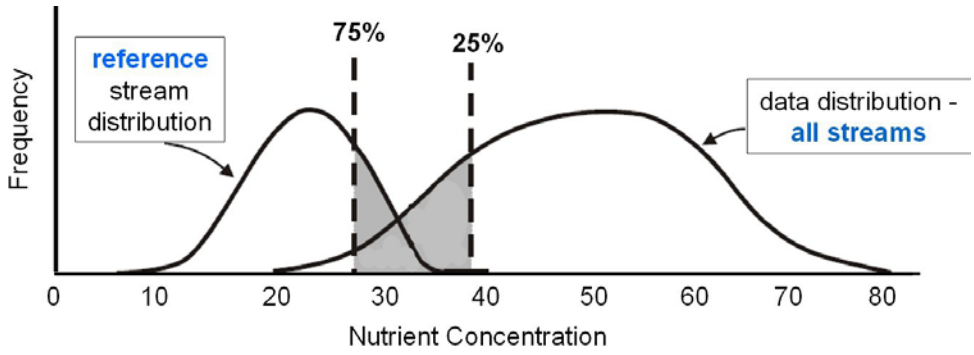


Figure 12. Generalized data distributions for reference waters versus all streams in an area of interest, here considering streams. The U.S. EPA (2000a,b) recommends considering the 75th percentile of the TN and TP data as reference conditions. Alternatively, if reference waters are not available, then the 25th percentile of data from all streams in the area could be used as the targeted conditions. The diagram shows that the 75th percentile of data for reference waters approximates "minimally impacted" conditions, and is similar to, although lower than, the 25th percentile of data from all streams. If in a highly urban or agricultural area, the 5th to 15th percentile of data should be used, rather than the 25th percentile of data.

the 25th percentile of data from all similar waterbodies within the ecoregion or sub-ecoregion of interest (Figure 12). If waters with reference conditions are not available, then the 25th percentile of the range of P and N concentrations from all waters (all lakes and reservoirs, or all rivers and streams, past ~10 years of data) can be used.

Importantly, the U.S. EPA recognized that in some areas, waters may be so polluted that even the lowest (best) 25% of the observations will represent highly polluted conditions. If most waters within a sub-ecoregion of interest are degraded − for example, in highly urbanized or agricultural areas − then the U.S. EPA (2000c) recommended taking the 5th to 15th percentile (5th percentile if the area has most land cover in urban or agricultural use) of the range of P and N concentrations as the targeted numeric P and N criteria. The range in percentiles allows some adjustment depending on the known severity of the water quality degradation. For example, in dense corn belt areas, the U.S. EPA recommended using the 5th percentile of the range of P and N concentrations, whereas in moderately urbanized areas the 10th or 15th percentiles could be used. As another alternative, reference TN and TP concentrations from an adjacent sub-ecoregion with similar characteristics but less human disturbance could be used as the numeric P and N criteria.

Note that the "reference approach" is confusingly named, since most reference streams (1st through 74th percentiles of data, including those with the most pristine conditions, come out as "better than natural" using the U.S. EPA (2000a-c) reasoning. Focus on the reference condition at the 25th percentile will fail to protect remaining pristine or highest-quality waters. "Minimally impacted" may not accurately describe the reference conditions either, because the 25th percentile of all-streams data may be a considerably more degraded condition than minimally impacted.[2]

The reference condition approach at least has the virtue that it is straightforward to use in freshwaters, as illustrated by the following examples:

Illinois River at Peoria, north-central Illinois (nutrient ecoregion VI, Corn Belt and Northern Great Plains; sub-ecoregion #54 - Figure 13): The 25th percentile of all-streams data (Table 3) are the numeric criteria suggested by the U.S. EPA (2000e). These highly polluted waterbodies have recommended criteria of ≤ 2,461 µg TN/L (calculated from TKN + NOx) and ≤ 72.5 µg TP/L using the U.S. EPA reference (minimally impacted) approach.

[2] In addition, the reference approach does not target protection of designated uses per se, as does the Clean Water Act. It is assumed that reference or minimally impacted conditions will protect designated uses, but that depends on the designated use. For example, many waterbodies have designated use by (generic) "aquatic life." Reference waters typically support sensitive fish species and are characterized by low algal biomass. In Missouri the desirable present-day condition for designated use by "aquatic life" has been clarified to target sportfish species (Missouri Department of Natural Resources 2016), which are tolerant of nutrient-enriched waters (e.g., Hoyer and Canfield 1996). The low algal biomass typical of reference conditions could be inadequate to support the designated use by aquatic life - sportfish species

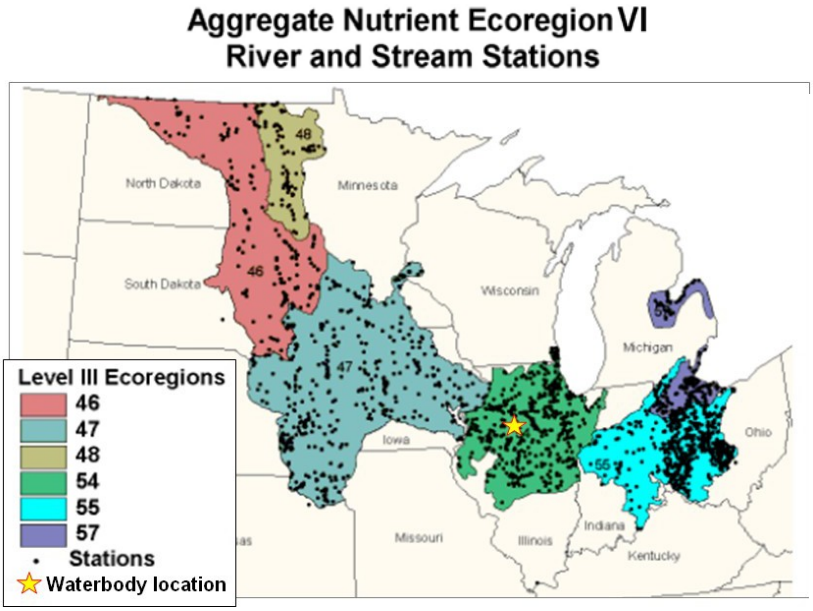


Figure 13. Map showing the locations of the nutrient sub-ecoregions within level III ecoregion VI (see Figure 11), also showing the location of the Illinois River at Peoria. Modified from the U.S. EPA (2000e).

Table 3. Numeric TN, TP, and chlorophyll a criteria recommended for the Illinois River at Peoria, using the statistical method (U.S. EPA 2000e).

Parameter	25th percentiles based on all seasons data for the decade*
	P25 - all seasons*
TKN	663 µg/L
NOx (NO_3^- + NO_2^-)	1798 µg/L
TN (calculated)	2461 µg/L
TN (reported)	2950 µg/L
TP	72.5 µg/L
Chlorophyll a (F)*	2 µg/L
Chlorophyll a (S)*	7.01 µg/L

*Suggested values were also given for turbidity. Values for chlorophyll *a* are corrected to exclude pigment from dead algal cells (F, fluorometric method; S, spectrophotometric method). Note that numbers based on more recent decadal data are not available.

Lake Chippewa, northwestern Wisconsin (nutrient ecoregion VIII, Nutrient-Poor Largely Glaciated Upper Midwest and Northeast, sub-ecoregion #50 - Figure 14, Table 4). This much cleaner waterbody has recommended criteria of ≤ 323 µg TN/L (calculated as above) and ≤ 9.69 µg TP/L using the U.S. EPA reference (minimally impacted) approach.

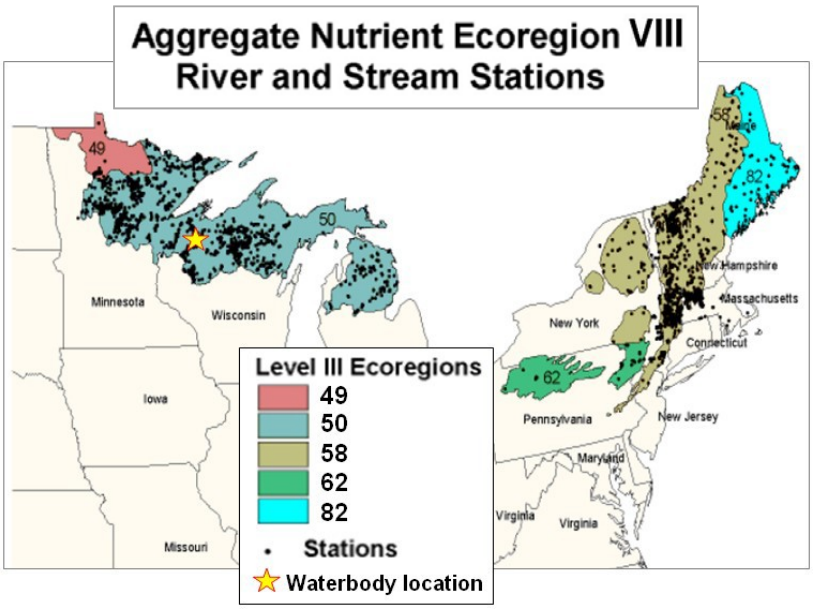


Figure 14. Map showing the locations of the nutrient sub-ecoregions within level III ecoregion VIII (see Figure 11), also showing the location of Lake Chippewa, WI. Modified from the U.S. EPA (2000f).

Table 4. Numeric TN, TP, and chlorophyll a criteria recommended for the Lake Chippewa, northwestern WI, using the statistical method (U.S. EPA 2000f).

Parameter	25th percentiles based on all seasons data for the decade*
	P25 - all seasons*
TKN	320 µg/L
NOx (NO_3^- + NO_2^-)	3 µg/L
TN (calculated)	323 µg/L
TN (reported)	400 µg/L
TP	9.69 µg/L
Chlorophyll *a* (F)*	1.38 µg/L
Chlorophyll *a* (S)*	2.46 µg/L

*Suggested values were also given for Secchi depth (4.2 meters). Values for chlorophyll *a* are corrected to exclude pigment from dead algal cells (F, fluorometric method; S, spectrophotometric method). Note that numbers based on more recent decadal data are not available.

To some extent, these differences may reflect differences in natural conditions, but to a large extent they also implicitly assign higher (weaker, i.e., less protective) criteria to waters that are polluted – for example, in Figure 13, reflective of the name of one of the ecoregions, the "Corn Belt." It is, then, somewhat like dividing students into two classes of good students and poor students, and grading each of the two classes on a curve set by the scores of each class.

2. Literature approach - This approach relies upon the published science literature as the source of numeric nutrient concentrations indicating minimally impacted (nutrient-poor or oligotrophic), moderately nutrient-enriched (mesotrophic), and highly nutrient-enriched (eutrophic or hypereutrophic) conditions (Tables 1 and 2). For example, resource managers might appropriately select mesotrophic levels of TN and TP to protect the designated uses of lakes that lie within watersheds containing naturally nutrient-rich soils. For smaller (wadeable) streams, numeric nutrient criteria could be based on cumulative frequency distributions of literature data (U.S. EPA 2000b) (e.g., Table 2). A margin of safety can be provided by adopting numeric criteria targets that are 5-10% more protective than the literature values. The load reductions needed to achieve the targets can be estimated using the dilution capacity of the streams at very low flows (e.g., the lowest 30-day average flow likely to be measured in 1 of 10 summers).

3. Stressor/response (or effects-based) approach, sometimes called empirical stressor-response modeling, is discussed at greatest length here because it has been favored by most states, although it requires large datasets for both water quality and biota (e.g., fish and macroinvertebrates including sensitive species) to derive numeric nutrient criteria, and can also require more advanced statistical expertise depending on the analysis. As explained above (Section II.b.2, pp. 24-26), recreational uses have also been considered in developing numeric nutrient criteria using stressor-response approaches, although some types of data are limited or have had limited utility (e.g., user perceptions).

The stressor/response approach has been subject to misuse (see Section VIa,d), but it can be sound if used as statisticians designed and intended. This approach relies on predictive relationships linked to protection of designated uses (e.g., trophic state classifications, models, recreation, biocriteria). The stressor/response approach can show a clearer link between nutrient criteria and designated uses than the reference approach, and it allows consideration of different management objectives (U.S. EPA 2000a).

As a caution, substantial data are usually required that extend beyond nutrient concentrations and response variables, including measurements of environmental factors that potentially can confound the estimated relationships, such as temperature or flow conditions (U.S. EPA 2010). The stressor-response approach can be used when sufficient data are available to estimate, accurately, a relationship between nutrient concentrations and a response measure that is directly or indirectly related to a designated use of the waterbody(s) - such as a biological index or a recreational use measure (U.S. EPA 2010). Numeric nutrient criteria that are protective of designated uses are derived from the estimated relationship. There is strong precedent for use of biological data to support stressor-response criteria derivation. A good example is given in Robertson et al. (2006, 2008) for Wisconsin streams, which suggested numeric concentrations that would minimize degradation of the biotic integrity of the waterbodies.

Stressor-response relationships can be valuable in characterizing progressive degradation in waterbodies with increasing nutrient pollution stressors (TN, TP). They help to identify the nutrient concentrations that support valued ecological attributes such as a healthy fish community (Stevenson et al. 2008). In this approach, nutrient concentrations are quantitatively related to valued ecological features (e.g., water clarity) and specific benefits (e.g., smaller, fewer noxious algal blooms) of protecting ecosystems at specific nutrient levels. Thus, stressor-response relationships can be used to establish numeric criteria at nutrient stressor concentrations that provide the desired level of protection for the valued ecological attribute (Stevenson et al. 2008). They can be assessed across various scales, from small laboratory bioassays to field surveys that correlate nutrient levels with abundance of desired aquatic biota or indices of aquatic community health.

In such efforts to quantify stressors, identification of thresholds can be especially valuable in justifying criteria at specific nutrient levels (Muradian 2001, Stevenson et al. 2008). A threshold is defined as a sudden change in desired conditions along a nutrient gradient. At or below the concentration threshold, the designated use(s) will be protected. The first threshold is commonly related by the analyst to the 75th percentile of data for sensitive species numbers (or abundances), that is, the concentration above which there is a sudden 25% decrease in species numbers. For example, TP concentrations below the threshold concentration at the 75th percentile of sensitive macroinvertebrate species would protect 75% of those species. At slightly higher TP concentrations, a sudden loss of ~25% of those species would occur.

The U.S. EPA (2010) published a recommended technical guide on "Using Stressor-response Relationships to Derive Numeric Nutrient Criteria," including four basic steps, briefly summarized as follows:

(1) Develop a conceptual model(s) for known relationships between N and P concentrations, biological responses, and attainment of designated uses for the specific study area and waterbody(s) of interest. A conceptual model is a visual representation of relationships among human activities, stressors such as TN and TP pollution, biological responses, and designated uses. Such models reflect accepted scientific knowledge and help guide numeric criteria development;

(2) Assemble data and conduct initial exploratory analyses, using variables (factors) that represent different concepts shown in the conceptual model (e.g., N and P concentrations, and response variables directly or indirectly related to designated uses, such as chla [noxious algal blooms] and DO [low DO, death of beneficial aquatic life such as fish]). Make simple graphs (e.g., scatter plots, bar graphs, box plots [box-and-whisker plots]) of the available data on nutrients and response variables in the waterbodies of interest, or maps of the data in the region of interest. Also conduct simple statistics such as correlation analyses to assess potential relationships among the variables (e.g., TP versus chlorophyll a);

(3) Assess stressor-response relationships between nutrient concentrations and selected response variables: For a group of similar waterbodies (e.g., lakes within a nutrient sub-ecoregion), conduct various analyses to assess whether causal and response variables are related.

(4) Derive numeric TN and TP criteria from these relationships, and evaluate the accuracy and precision of the estimated stressor-response relationships with documented analyses.

Some commonly used techniques in stressor-response analysis and actual derivation of the criteria are described below with supporting illustrations.

3a. Empirical Models and Assumptions that Must Be Met for Their Use

Empirical models are useful when adequate data allow development of statistically significant, strong relationships. The parametric statistical tests used in these models have the underlying assumption that the data are normally distributed – that is, the data follow a bell-shaped curve. Normal distributions are defined by two parameters, the mean and the standard deviation. In a classic normal distribution, 68% of the area under the bell curve is within one standard deviation of the mean, and ~95% of the area is within two standard deviations of the mean (Sokal and Rolff 2012).

Empirical models have been valuable in predicting algal biomass responses, for example, in lakes (Vollenweider 1968), large rivers (Basu and Pick 1996), and wadeable streams (Dodds et al. 1997, Biggs 2000, Smith et al. 2013). Most commonly, empirical regression models have been used to quantify relationships between nutrients and algal biomass or other biological metrics, considering data from a group of similar waterbodies (e.g., Jones and Bachmann 1976, Dodds et al. 1997). They have been used to set the general parameter boundaries (TP, TN, chlorophyll a, water transparency as Secchi depth) for lakes (Wetzel 2001 and references therein). Suplee et al. (2015) reviewed empirical modeling methods for numeric nutrient criteria applied to streams and rivers, and concluded that over the past ~15 years, there has been growing consensus about nutrient thresholds and responses (e.g., Dodds et al. 1997, Biggs 2000, Dodds and Welch 2000, U.S. EPA 2000a, Sheeder and Evans 2004, Stevenson et al. 2006, Wang et al. 2007, U.S. EPA 2010, Morgan and Kline 2011, Chambers et al. 2012, Stevenson et al. 2012, Smith et al. 2013, Smucker et al. 2013, Evans-White et al. 2014). Suplee et al. (2015) concluded, from these works, that depending on the waterbodies, 210-1,700 µg TN/L and 10-90 µg TP/L have been recommended to minimize noxious algal growth, DO deficits, and/or undesirable changes in valued biological communities such as fish and macroinvertebrates.

3a.1. Simple linear regression (SLR)[3]

Considering the most often-used empirical model, SLR, initial data analysis (guided by scatter plots and other exploratory information gathered in step 2 above) provides an estimate of the extent to which (if any) there is a linear relationship between a causal variable and a response variable. A stressor-response relationship estimated by SLR predicts the value of the response variable at a given nutrient concentration, so if the desired value of the response

[3] "Simple" refers to the fact that the models are two-dimensional, one causal variable and one response variable are the focus, and the simplest relationship is considered to be linear. The relationships in these models are mainly linear but may deviate from linearity over portions of the data range; a relationship can be linear over some range, then hyperbolic over another, then flat over another.

variable supportive of the designated use of the waterbody(s) is known (e.g., mesotrophic lake conditions at 20 µg chla/L, based on what is known about the system and/or on published science such as the information summarized in Tables 1 and 2), then the stressor-response relationship can indicate the average TP level corresponding to that chlorophyll concentration.

SLR summarizes the conditional mean (average) relationship. Several major assumptions must be met before a linear regression analysis can be accepted. First, the analyst fundamentally must determine that a straight line is appropriate for representing the relationship being modeled. This assumption can be evaluated statistically, but often is checked based on visual inspection of the data. The examples below show (Figure 15) appropriately linear data, versus (Figure 16) inappropriate application of simple linear regression analysis to the data.

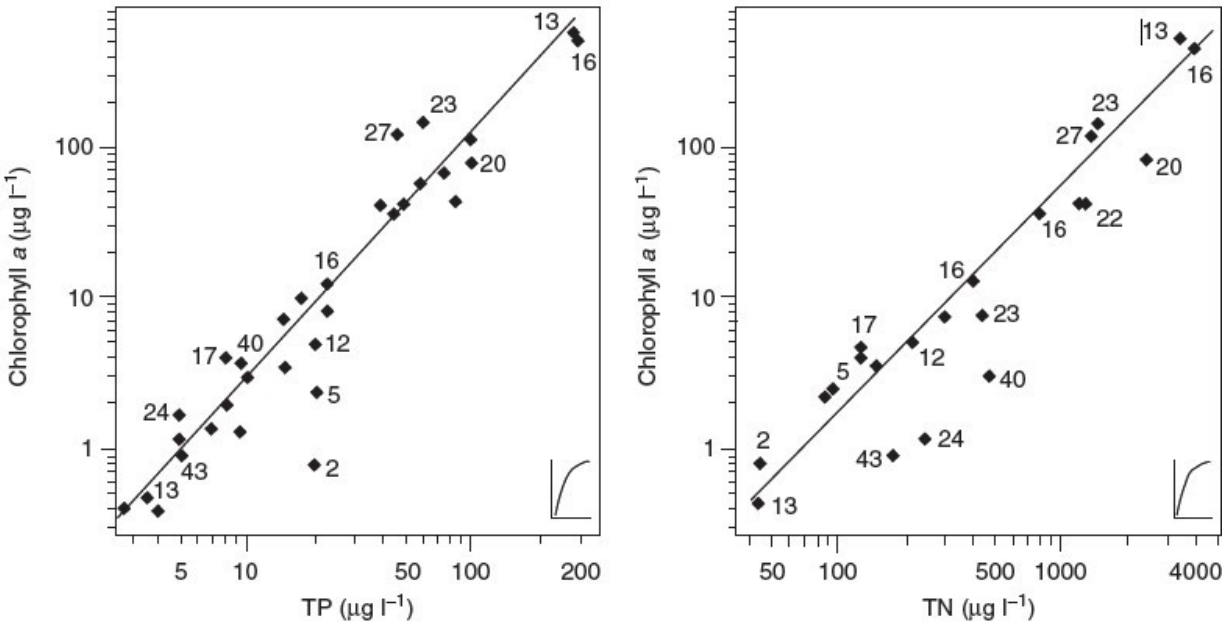


Figure 15. Linear relationship between chlorophyll a and (a) TP or (b) TN concentrations in surface and near-surface waters of a set of temperate-zone lakes in summer (May, June, July). Numbers next to points represent the N:P ratio (by mass). Note that lakes with exceptionally low and exceptionally high N:P ratios are outliers in the TP-chlorophyll a and TN-chlorophyll a relationships, respectively. Modified from Sakamoto (1966) in Kalff (2002).

In Figure 15, simple visual inspection indicates that a SLR would fit the data well. This figure depicts a major advance in the science of eutrophication, the development of statistically significant linear relationships between phytoplankton biomass in lake surface waters and water-column concentrations of TP and/or TN (Sakamoto 1966, Dillon and Rigler 1974, Jones and Bachmann 1976, Cooke et al. 2005). Based on simple models developed from these relationships (e.g., Dillon and Rigler 1975), quantitative estimates of in-lake nutrient concentrations were predicted from watershed nutrient inputs (e.g., the number of cottages around a lake). Note that for highly variable data, in this case comparing lakes from variable regions of the world, the data can be plotted on a log-log scale. Use of the log of the N or P

data reduces the influence of extreme outlier values (that is, very high or low N or P concentrations) and facilitates detection of a linear relationship if present.

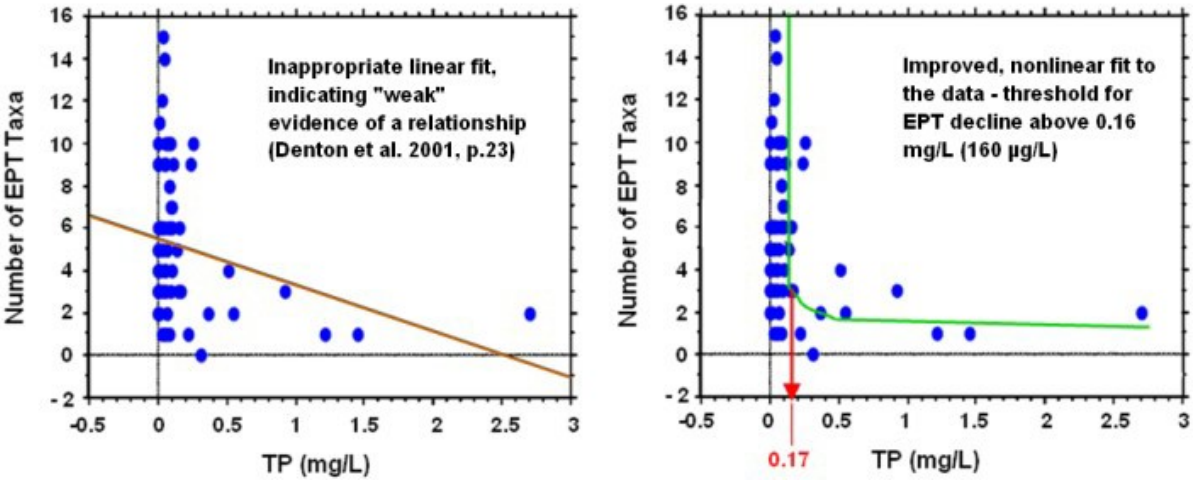


Figure 16. Relationship between TP concentration and the EPT metric (= sensitive aquatic macroinvertebrate insects including ephemeropterans - mayflies, plecopterans - stoneflies, and trichopterans - caddisflies): Left – Inappropriate fit of the data to a straight line resulted in misinterpretation as "no strong relationship" between TP and sensitive macroinvertebrate taxa (Denton et al. 2001). Right – The appropriate fit to the data showed a strong threshold response of the biota (e.g. Stevenson et al. 2008 and references therein) to increasing TP pollution, with severe impacts at TP concentrations exceeding ~170 µg TP/L.

In contrast, Figure 16 (left panel) illustrates a failed attempt to assess whether TP concentrations had a significant, major impact on sensitive stream macroinvertebrates in a nutrient sub-ecoregion. Visually, the data did not fit a line - instead (right panel), the data followed a classic threshold response of sensitive biota to low levels of a stressor (e.g., Stevenson et al. 2008), wherein most sensitive taxa occurred at relatively low TP concentrations (here, < 160 µg TP/L), and the number of sensitive taxa declined dramatically above 170 µg TP/L.[4] The resource managers inappropriately "forced" a linear analysis on non-linear data, and missed a clear biological response threshold. Importantly as well, the resource managers had rejected the U.S. EPA's (2000g) recommended numeric nutrient criterion of 170 µg TP/L for streams within that nutrient sub-ecoregion. Yet, the actual data for the response of sensitive stream macroinvertebrates to increasing TP pollution strongly supported the U.S. EPA recommendation.

The second assumption that must be met for appropriate use of SLR in stressor-response evaluations is that the data and residual values must be normally distributed (also referred to as parametric statistics). A residual is defined as the difference between the mean value (modeled as the line in a SLR) and the observed value for each data point (U.S. EPA 2010). A normal distribution refers to a dataset for a variable such as TP that, when plotted, has a bell shape wherein most of the data occur around the middle (average) of the TP concentration

[4] These relatively high numbers (160-170 µg TP/L, reflect the fact that the stream drains a geologically P-rich area.

bell shape wherein most of the data occur around the middle (average) of the TP concentration range, with symmetrically tapering "tails" of lower and higher TP values to the left and right of the middle. Analogously, a bell-shaped curve describes the grades for a hypothetical class of students: most students earn a middle (B^-, C, or D^+) grade – the middle of the curve, whereas few students receive a high grade (A to B^+) or a low grade (D or lower) within the "tails" on either side of the bell.

Many water quality, total taxa, and relative biota abundance data often are not normally distributed. If a normal distribution is lacking, the data can be transformed (log transformation, arcsine-square root transformation, etc.) to make them normally distributed. As a second requirement in addition to data that are normally distributed, the data residuals must also be normally distributed; and a third requirement that must be met for use of SLR is that the variance across the residuals is constant - that is, the scatter of the residual values about a line of the predicted values is constant across the range of residuals. If these assumptions are not met, the analyst may mistakenly conclude that the data are strongly, linearly related when, in reality, that conclusion is not statistically supported. Departures from linearity can also result in underestimating, or even completely discounting, an important independent variable that strongly influenced the dependent variable (Robertson et al. 2006).

3a.2. Multiple Regression Analysis (MR)

As mentioned, SLR targets the average relationship between one stressor and one response variable. Resource managers are sometimes interested in describing stressor-response relationships at different points in the response variable range, such as the response of sensitive macroinvertebrate metrics at the high end of the TP concentration range. Multiple regression analysis (MR) extends SLR, with the same required assumptions, to provide an estimate of the linear relationships between a response variable and two or more independent variables - that is, each causal (explanatory) variable is assumed to act independently of effects of the other causal variable(s). MR enables modeling of the effects of several different stressors on the response variable simultaneously – for example, the effects of TN, TP, and temperature on suspended microalgal chla.

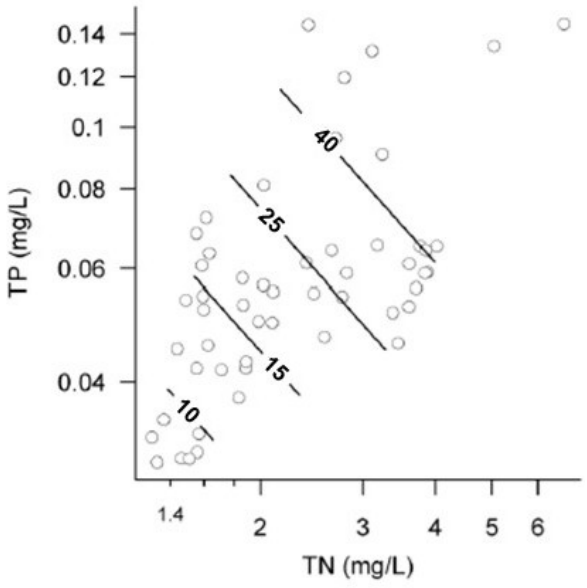


Figure 17. The relationship between TP, TN, and chla in a lake of interest, modeled using MRA. The plotted circles indicate combinations of TN and TP values observed in the data. Contour lines indicate modeled mean chla concentrations (μg/L) associated with particular combinations of TN and TP. Modified from an example given in U.S. EPA (2010).

The example in Figure 17 concerns a lake where both TN and TP are statistically significant predictors of chla, so that both TN and TP criteria should be specified to achieve a targeted chla concentration. Here, to maintain an average chla concentration

of 10 µg/L, criteria could be set at 1.6 mg TN/L and a corresponding TP level of 0.035 µg/L (U.S. EPA 2010). Following the graph, a lower TN criterion would require a higher TP criterion to maintain the same average chla concentration.

For example, in a set of streams and rivers, Biggs (2000) used MR models that combined dissolved inorganic N, dissolved inorganic P, and the number of days of accrual to assess influences on benthic algal biomass as chla. The models explained ~45% of the variation in mean monthly benthic chlorophyll, and ~73% of the variation in maximum benthic chlorophyll among sites. In streams with Infrequent floods and long accrual periods (more than 100 days), but not in more flood-prone streams, a relatively small increase in dissolved inorganic N and P significantly increased the frequency of high benthic algal biomass events.

Although both TN and TP have been shown to be important in controlling algal/plant growth in many aquatic ecosystems (Dodds and Smith 2016, and references therein), over a particular range of interest, there can be a stronger relationship between algal/plant growth and one nutrient or the other. The more strongly related nutrient is often emphasized in developing stressor-response relationships, especially if more data are available for that nutrient. Another consideration is that some causal variables, such as TN and TP, tend to be strongly correlated, which can confound interpretations about the separate influence of each and can greatly increase the uncertainty of model estimates (Kutner et al. 2004). MR is useful in providing information about controlling influences of independent causal variables such as temperature, light, flow, and grazing activity on biota response across nutrient gradients.

3b. Useful analyses when the required assumptions of SLR and MR cannot be met

Importantly, many biota responses to changes in nutrient concentrations are nonlinear. The assumptions required for SLR are, in various situations, too restrictive to enable accurate modeling of the observed data (U.S. EPA 2010). Other, nonparametric modeling approaches should be used when residuals are not normally distributed with constant variance (see p.38 for definitions of these terms). The term “nonparametric” refers to statistical techniques that do not require the data to follow a normal distribution. Nonparametric tests are sometimes called distribution-free tests because they do not assume any specific distribution. They can be applied when the major relationship between the stressor and the response is not linear.

Various statistical methods are available for identifying nonlinear and threshold relationships (Dodds et al. 2010). The examples given below are methods that have been used to develop numeric nutrient criteria for waters in the Mississippi River basin. Differences in the criteria suggested from these examples depend on the groups of waterbodies, the variables considered, the extent to which reference or minimally impacted waters were included in the analysis, and choices made by the user in running the automated statistical programs. More about statistical threshold techniques is given in Appendix 1.

3.b.1. Quantile Regression Analysis (QR), including Additive QR Smoothing (AQRS)

Quantiles are defined as any set of regularly spaced intervals in a set of data that are ordered from the lowest to the highest values (e.g., 10% quantiles in a dataset, or quartiles for 4 intervals in a dataset, or 100 intervals [percentiles] in a dataset) (U.S. EPA 2010). QR measures the

median condition (50th percentile) of a response variable and estimates multiple rates of change of the response variable along a stressor gradient (Cade and Noon 2003, Koenker 2005, Caskey et al. 2012) (Figures 18 and 19). It also enables consideration of various quantiles, such as the TP concentration protective of 75% [75th percentile] of macroinvertebrate species in a set of similar streams. QR estimates are more robust than those of SLR against outliers (extreme values) in the response data, and it is better suited to the "wedge-shaped data" that are often characteristic of biological responses to nutrient stressors (Bryce et al. 2008).

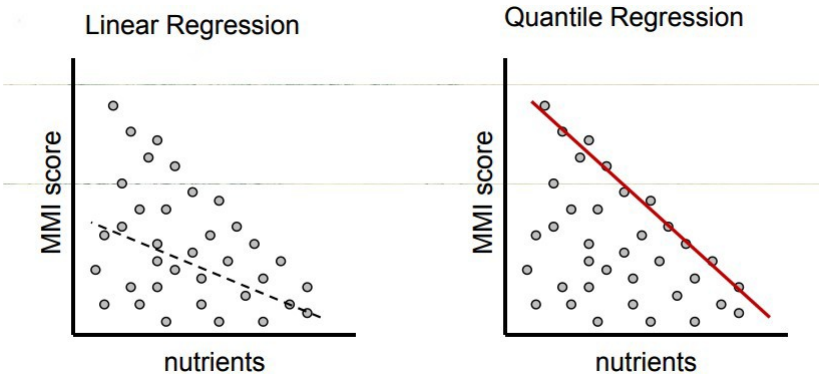


Figure 18. Diagram depicting the advantage of QR over SLR in assessing stressor-response relationships for the characteristic "wedge-shaped" data for responding biota. MMI - multi-metric index to assess macroinvertebrate community health. From Canton et al. (2012).

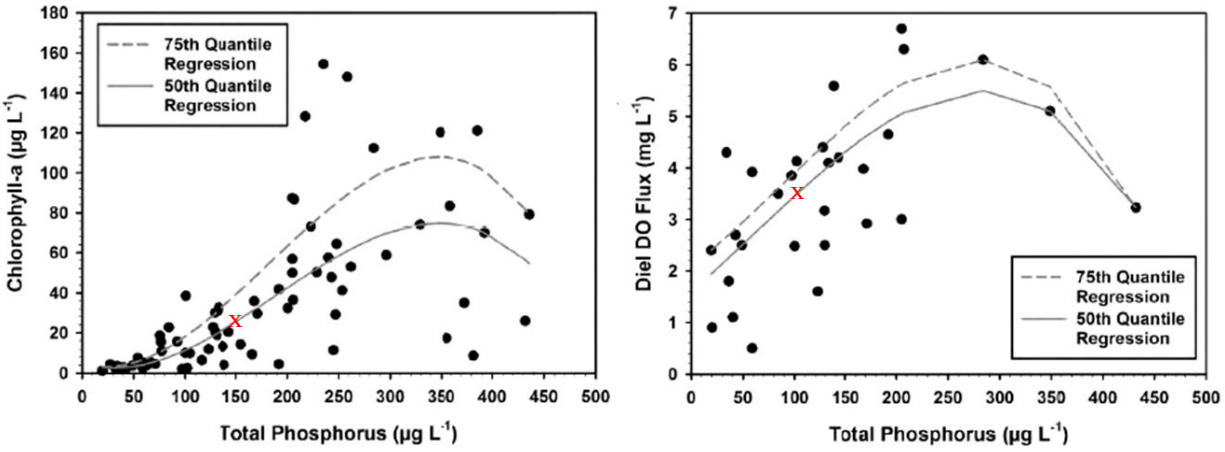


Figure 19. Relationship between TP and (left) chlorophyll a or (right) diel DO change (flux) in Minnesota rivers, analyzed using QR. The regression lines were based on both nonwadeable and wadeable streams using 50th and 75th percentile QR spline fits (R Development Core Team 2009). The first graph indicates that the median TP (50th quantile line) should be ≤ ~150 µg/L (red X) to keep chla ≤ ~25 µg/L; the second graph indicates that median TP (50th quantile line) should be ≤ ~80 µg/L (red X) to keep diel DO flux ≤ ~3.5 mg/L. Note that in both graphs at highest TP levels, the lines curve downward. This decreasing chla at highest TP concentrations suggests that another factor(s) that co-occurred with the high TP caused the decline, such as light limitation due to self-shading. The decrease in diel DO flux at highest TP levels, similarly, may have resulted from light limitation or other co-occurring factor(s), such as elimination of some taxa during night-time DO sags under extremely high TP conditions. From Heiskary et al. (2013).

In rivers and streams, the relationships between nutrients, algal biomass, and biota at higher trophic levels such as fish often appear to be weak, masked by complex interactions. Usually, all of the factors that affect the ecological process of interest (e.g, nutrient effects on algal

biomass) are not measured and included in the statistical models used to assess relationships among variables (Cade and Noon 2003). This failure to include all of the relevant variables occurs because of time or funding constraints, or insufficient knowledge or ability to make the necessary measurements. Thus, in one part of the response variable distribution (range), there may be only a weak relationship, or no predictive relationship, between the mean of the response variable distribution and the measured predictive factors – whereas, in another part of the response variable distribution, however, there may be stronger, useful predictive relationships with the response variable (Cade and Noon 2003).

The need to understand significant ecological relationships in such complex environmental settings has made QR useful in stressor-response studies (Caskey et al. 2012). QR provides a way to estimate the quantiles of a response variable distribution that enables a more complete view of possible causal relationships between variables. In other words, QR can be used to identify nutrient threshold concentrations that protect the designated uses of waterbodies, based on analysis of field data for biota that have complex relationships with nutrients depending on various physical/chemical conditions (Heiskary et al. 2013). The wedge-shaped data plots reflect the limitation of biological attributes (for example, species richness) by nutrients on the outer or upper edge of the wedge (Bryce et al. 2008). Inside the wedge, the biological response is limited by other, unmeasured environmental factors. Due to the influence of these other factors, there is unequal variation in the biological attribute at different nutrient levels. Unequal variation implies that there is more than one rate of change describing the relationship between a response variable (e.g., suspended algal chlorophyll a) and a predictor (causal) variable (e.g., TN or TP). This unequal variation can make field data inappropriate for analysis by SLR or related techniques that assume equal variance across the data.

QR estimates multiple rates of change from the minimum to the maximum response, thus providing a more complete description of the relationships between variables than other regression methods (Cade and Noon 2003). QR can be used when the assumptions required for SLR are not applicable. A related statistical technique, AQRS (Figure 20), is similar to linear QR, except that instead of fitting a single line to the data, AQRS fits a regression line to subsets of the data (e.g, Figures 20 and 21). Thus, AQRS can also be used to identify thresholds (breakpoints or changepoints) in addition to fitting the outside of the data wedge.

As shown by AQRS (Figures 20-22), and as noted above (p. 29), thresholds (breakpoints) in response data occur wherein a small change in nutrient concentrations corresponds to a relatively large change in the biotic communities. Thus, the thresholds for many of the biota responses generally are only slightly higher than the reference concentrations (e.g., Robertson et al. 2008). If relationships between nutrient concentrations and biotic integrity are used to define criteria, the final nutrient criteria should be selected so as to minimize degradation in the biotic integrity of the waterbody (Robertson et al. 2006).

AQRS has sometimes shown stronger predictive stressor-response relationships than other available techniques such as nonparametric changepoint analysis (nCPA, below). The

"smoothing" (S) refers to smoothing techniques such as weighted scatterplot smoothing (LOWESS smooth), a nonparametric tool in regression analysis that creates a smooth line through a data plot (U.S. EPA 2010). It is usually used with highly variable ("noisy") data or sparse data, or when weak relationships obscure any line of best fit. Nonparametric smoothing techniques can be used to identify the "true signal" hidden in noisy data. Because some type of distribution (e.g., normal distribution) is assumed in advance, parametric fitting of lines or curves to the data can misrepresent the data. In such cases, nonparametric smoothing would be a better approach.

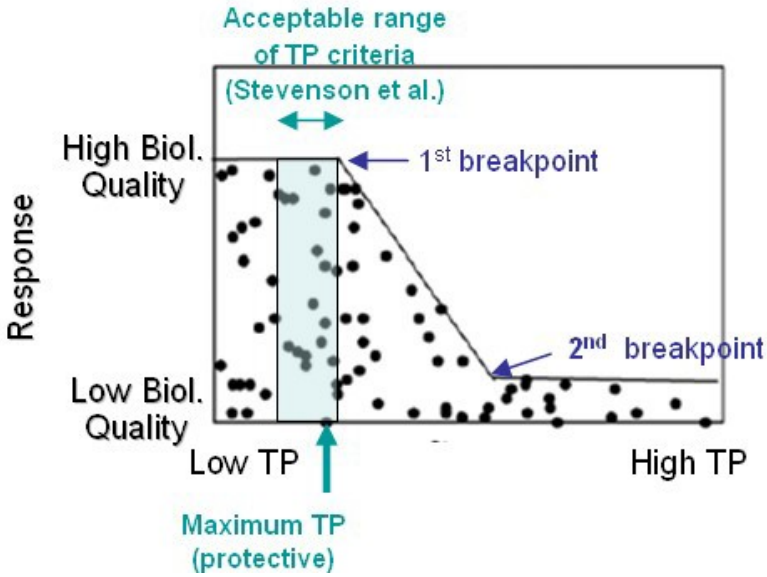


Figure 20. Example of use of additive quantile regression smoothing analysis (AQRS) to derive a protective numeric TP criterion. Here, "response" (Y axis) refers to the response of sensitive biota to nutrient pollution. See Wang et al. (2007), Brenden et al. (2008), and Stevenson et al. (2008).

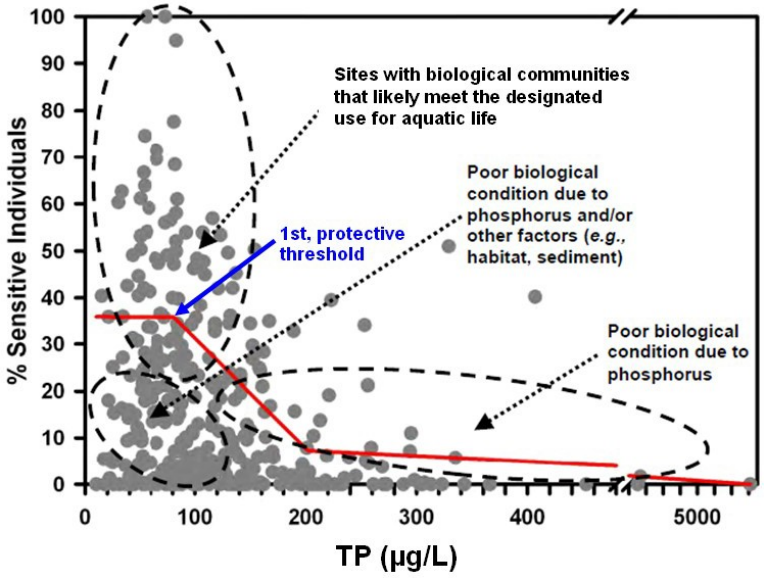


Figure 21. Relationship between TP and the percent of sensitive fish in central Minnesota streams, as determined using additive QR smoothing (AQRS, red line). This diagram also shows the typical wedge-shaped data to which QR is suited. From Heiskary et al. (2013). As shown in Figures 21 and 22, QR applied to biological field data can reveal more than one breakpoint. From Heiskary et al. (2013).

Smoothing techniques such as LOWESS find a curve of best fit to the data without assuming any distribution shape. They enable a flexible approach in representing data and are easy to use, but they cannot be used to obtain a simple predictive equation for a dataset and may require

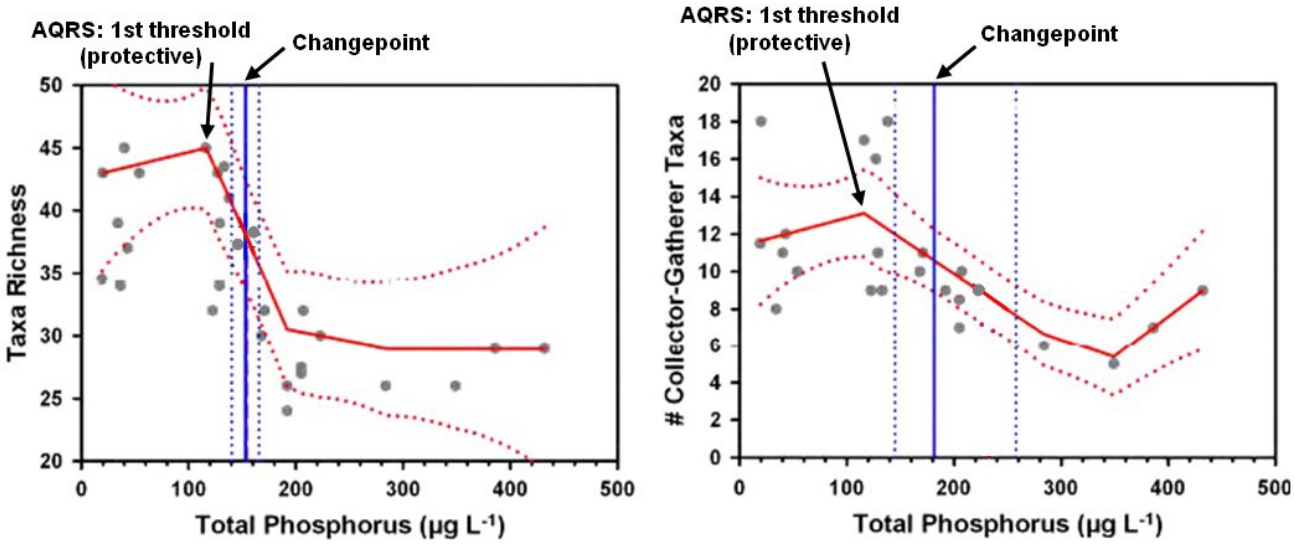


Figure 22. Relationships between TP and macroinvertebrate metrics for all Minnesota streams statewide (red line = additive quantile regression with 90% confidence bands, blue line = changepoint analysis with 90% confidence bands). In both examples, AQRS indicated a more protective TP threshold than changepoint analysis. Left, taxa richness: The TP threshold was ~110 µg TP/L using AQRS, whereas changepoint analysis indicated a breakpoint at 127 µg TP/L. Right, number of collector-gatherer taxa: AQRS indicated a TP threshold of ~110 µg TP/L, whereas the changepoint was 182 µg TP/L. Modified from Heiskary et al. (2013).

substantial guesswork on the part of the analyst. The solid red lines in Figures 21 and 22, which are lines through data plots that had no apparent pattern, were drawn using LOWESS. Peer-reviewed science contains many other examples of LOWESS curve fitting, such as Dodds et al.'s (2002) analysis, wherein data for seasonal means of benthic algal biomass (as chlorophyll), nutrient concentrations, and other environmental data were collected for nearly 300 sampling periods in temperate streams (1st dataset) and from a subset of 620 stations in the USGS National Stream Water-Quality Monitoring Networks. In that study, the greatest portion of variance in mean and maximum benthic stream algal biomass (~40%) was explained by TN and TP concentrations. Statistically significant breakpoints occurred at ~40 µg TN/L and 30 µg TP/L, above which mean benthic algal biomass substantially increased.

Automated programs are available for analysis of data using AQRS (e.g., Koenker 2016). This powerful tool has several caveats: AQRS requires substantial data - ideally, paired data - to work well. For example, the biological data for sensitive species shown in Figures 21 and 22 should have been collected at a suitable lag period after the TP data were collected - but not months or years before or after the TP data were collected. Interpretations may still be limited since correlation does not necessarily mean causation; in addition, if N and P concentrations are significantly correlated (which is common), their effects cannot be separated. AQRS is described here as an example of threshold statistical techniques. Other examples are provided in Appendix 1.

4. Mechanistic (process-based, predictive, or mathematical) modeling is under-utilized because process-based computer models are complex and require (i.e., should only be conducted by) specialists with modeling expertise (U.S. EPA 2009a). This approach is applicable for more complex aquatic systems such as large rivers, where inadequate empirical data are available or somewhat unique confounding factors affect the eutrophication response data are available or somewhat unique confounding factors affect the eutrophication response (Carleton et al. 2009, Bierman et al. 2013, Flynn et al. 2015). Mechanism-oriented (process-based) models allow simultaneous evaluation of multiple response endpoints (e.g., algal data are available or somewhat unique confounding factors affect the eutrophication response

(Carleton et al. 2009, Bierman et al. 2013, Flynn et al. 2015). Mechanism-oriented (process-based) models allow simultaneous evaluation of multiple response endpoints (e.g., algal biomass, DO, pH, supersaturation); they also enable predictions outside of the original calibration range, and they can integrate site-specific factors such as point sources, turbidity, and season into a unified water quality management framework (Flynn et al. 2015 and references therein). These models are used to describe and predict relationships between nutrient availability, algal uptake kinetics, light, flow, temperature, and other variables that link nutrient concentrations and loads to ecological responses. The monitoring, assessment, and modeling work should target the critically important period for nutrient control in temperate rivers, usually summer or late summer-early fall, when beneficial uses are most threatened by nutrient enrichment and noxious conditions are most common (Flynn et al. 2015). Data collection, model development, and analysis procedures should be structured to meet peer-review requirements so as to help ensure high quality prior to use of the modeling results (protective numeric nutrient criteria) by resource managers.

This more complex modeling differs from empirical modeling (see stressor/response approach below), a general term referring to any type of (computer) modeling based on actual data (empirical) observations alone. Use of both types of models can be helpful in developing comprehensive nutrient management for complex waters such as large rivers. The first study that used the mechanistic modeling approach to derive numeric nutrient criteria for large U.S. rivers was only recently published (Flynn et al. 2015, Suplee et al. 2015). Large rivers were defined as perennial waters that are unwadeable in the summer-early fall base flow period, with base flow annual discharge of ~40 cubic meters per second (m^3/sec, or ~1,413 cubic feet per second [cfs]), or base flow depth exceeding 1 meter. Process-based modeling approaches emphasized the Yellowstone River and major tributaries. A generalized model-based approach was developed that can be adapted for use in other large rivers and, more broadly, in other aquatic ecosystems (Figure 23). As a first step, ecological response variables that could be easily measured, had distinct thresholds of protection, and were defensible in supporting designated uses) were identified for the waterbody of interest (e.g., DO, pH - Bierman et al. 2013). Second, candidate models were identified and screened for appropriate characteristics (see Bierman et al. 2013) - for example, the capacity of a model to treat stratified systems differently than well-mixed systems. After tentative selection of models to be applied to the Montana rivers, Flynn et al. (2015) conducted field reconnaissance (e.g., dye tracer studies to evaluate transport; aerial photos used to assess bankfull channel width and streamside vegetation shading) for an entire year prior to their main sampling to assess whether the system behavior (e.g., travel time, stratification, mixing) matched the model features.

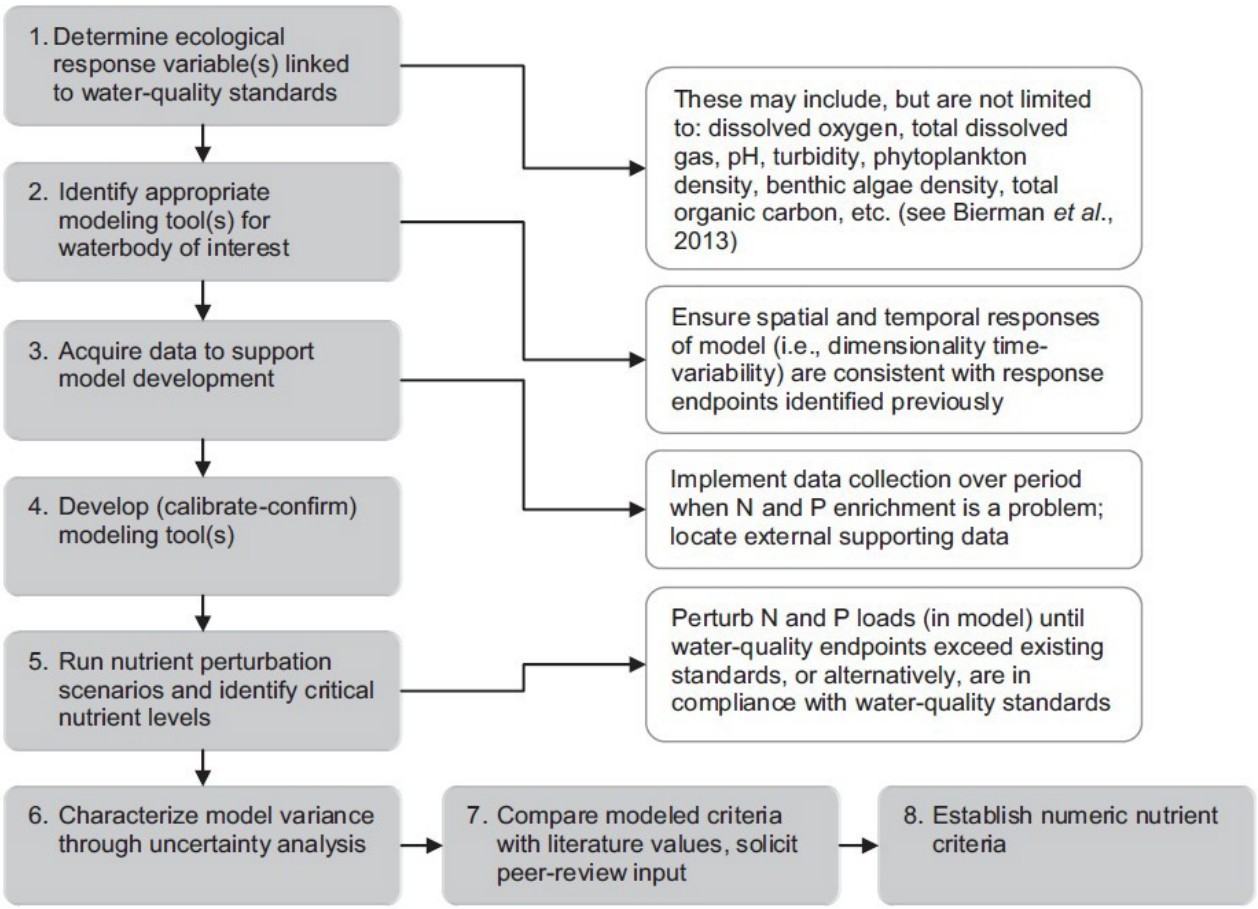


Figure 23. Approach for mechanistic (process) model-based numeric nutrient criteria in large rivers. This generalized methodology could be extended to other water bodies (e.g., lakes, reservoirs, estuaries, etc.). From Flynn et al. (2015).

Mechanistic models should be constructed (parameterized) with sufficient actual data from the system being modeled (but see Section VI). In the Montana rivers study, synoptic river surveys were conducted over two 10-day periods during the critical period for nutrient control over a 233-kilometer-long (~145-mile-long) river length. Sampling included eight mainstem river sites and 12 or more tributaries and irrigation return flows, which were characterized to constrain model calibration. Features considered included major inflowing tributaries, water withdrawals, wastewater treatment plant discharges, locations where nuisance conditions (e.g., algal blooms) could occur, climatic factors, and channel shifts (slope or morphology), following Mills et al. (1986) and Barnwell et al. (2004). These features reflected important boundary conditions or critical response areas in the model. Mainstem sites were located upstream from major tributaries to satisfy model assumptions about complete mixing. At each mainstem site, a vertically integrated water sample (including the total depth) was collected at 11 equidistant points along a transect (note - tributaries were grab-sampled). Benthic and suspended algal biomass as chlorophyll a was also sampled at the 11 points. Water chemistry measurements included total suspended solids, volatile suspended solids, TN, NOx [nitrate + nitrite], total ammonia, TP, soluble reactive phosphorus [SRP], carbonaceous 5-day biochemical oxygen demand, alkalinity, carbon-to-nitrogen-to phosphorus ratios, and benthic

and suspended algal biomass as chlorophyll a. Diel DO, pH, conductivity, and temperature, and flow, meteorology, river productivity, and respiration also were measured.

Numeric nutrient criteria were derived for the eight Montana rivers by simulating nutrient additions in the calibrated models to assess the level at which nutrients began to affect beneficial uses under low-flow conditions (Suplee et al. 2015). Either N or P was considered to be limiting; this model did not address co-limitation. Initial (baseline) nutrient concentrations were based on low-flow water quality data, or reference data for streams of similar size, or review of minimally impacted large rivers worldwide. Existing point sources were removed from the model simulations; otherwise, less impacted areas that were more sensitive to nutrient pollution would have been under-emphasized. Incremental increases in TN or TP concentration were assessed considering their modeled impact on predefined thresholds for each variable (e.g., minimum DO, maximum pH). The first variable exceeding a threshold was used to set the numeric nutrient criteria.

Importantly, for both N and P, the algal response to enrichment was nonlinear: at first, incremental increases in the nutrient concentration caused a large change in the algal response, before reaching a point where algal response (sensitivity) declined (as in Figure 9 above). That inflection point was close to where the numeric criteria were recommended. As would be expected given differences in shading, transport, flow etc., the Yellowstone river had different sensitivities to nutrients in the upstream versus downstream segments; pH was the most sensitive variable in the upstream reach, whereas algal biomass was the most sensitive variable in the downstream reach. Mechanistic model-based criteria for the Yellowstone River were 55 µg TP/L and 655 µg TN/L in the upstream area (between the Bighorn and Powder River confluences), and 95 µg TP/L and 815 µg TN/L in the downstream area (from the Powder River confluence to the Montana state line) (Suplee et al. 2015).

The Montana large rivers modelers noted some data needs for mechanistic models in setting numeric nutrient criteria, especially research on algal physiology, and extension of the models to enable reliable predictions for higher trophic levels such as fish and macroinvertebrates. They also cautioned against applying the results for a single river reach to segments much farther downstream that were not modeled. In addition, they recognized the need for alternative modeling approaches when considering waterbodies heavily influenced by nutrient pollution so that one nutrient is in extremely high abundance, or both N and P co-limit algal/plant growth.

5. Multiple Lines of Evidence (MLE; or weight of evidence) – The “hybrid” approach includes more than one of the above approaches, with the aim of strengthening support for the selected numeric criteria (U.S. EPA 2000c, 2010). It has been used by some states to develop numeric nutrient criteria for wadeable streams and rivers in New York (Smith and Tran 2010, Smith et al. 2013), rivers in Minnesota (Heiskary et al. 2013, Heiskary and Bouchard 2015), and nonwadeable rivers in Wisconsin (Weigel and Robertson 2007).

A weight-of-evidence approach was used by the state of Wisconsin to develop TP criteria for rivers and streams. It was based on a strong dataset analyzed by specialists knowledgeable about nutrient-related water quality/impacts in lakes, reservoirs, rivers, and streams, and also based on appropriate statistics for analysis of both water chemistry and biota data (Robertson et

al. 2006, 2008; Weigel and Robertson 2007). In that study, 34 nonwadeable rivers (41 sites) were sampled within U.S. EPA level III nutrient ecoregions VII (Mostly Glaciated Dairy Region) and VIII (Nutrient Poor, Largely Glaciated Upper Midwest) (Weigel and Robertson 2007; also see Appendix 1). An index of biotic integrity (IBI, consisting of ten component measures) was computed for fish assemblages, and 12 metrics were used to assess macroinvertebrates. The biotic measures ranged from poor to excellent in status. The reference (minimally impacted) conditions reported by the U.S. EPA (2000h) for sub-ecoregions in ecoregion VII ranged from ~21 to 80 µg TP/L. The reference conditions for sub-ecoregions in ecoregion VIII ranged from 6 to 40 µg TP/L (U.S. EPA 2000f). A type of statistical threshold analysis, regression tree analysis (Appendix 1), of TP on the suite of biological metrics identified an overall breakpoint (threshold) for TP at ~60 µg/L, above which fish and macroinvertebrate assemblages were consistently impaired in streams of both ecoregions.

Thus, the analysis included two approaches, Reference Condition and Stressor-Response (using two or more different threshold statistical techniques) including assessment of both fish and macroinvertebrates. The thresholds yielded from the Stressor-Response approach were within the range of minimally impacted conditions for sub-ecoregions within ecoregion VII, although somewhat higher than the range of minimally impacted conditions for sub-ecoregions within ecoregion VIII. It should be noted, despite the sound science reflected in the approach, that the greatest protection of sensitive biological metrics would have favored use of the lowest TP and TN breakpoints in setting numeric criteria. Wisconsin elected to impose moderate protection: Based on these and other data, a numeric TP criterion for the state's nonwadeable rivers was set at the midpoint of the TP range, 100 µg/L (Wisconsin Department of Natural Resources 2017).

IV.d. Effects-based criteria – "drawing the line"

For effects-based approaches when a clear threshold(s) in a desirable metric is identified by statistical analysis, the statistical programs can be used in line-drawing to identify the first threshold (e.g., Wang et al. 2007, Brenden et al. 2008). The program typically is modified to identify the threshold nutrient concentration that maintains 75% of the desirable metric, which can be used as the numeric nutrient criterion not to be exceeded. To be more protective, the numeric criterion should be set 5-10% lower than the indicated threshold nutrient concentration to include a margin of safety, following Stevenson et al.'s (2008) cautionary guidance (see Figure 20).

In contrast, when the data for a metric of interest gradually change across a nutrient concentration range, and/or when a clear, protective threshold in a response metric is lacking, then selection of protective numeric nutrient criteria requires a line-drawing procedure informed by the desired conditions for the metric (see examples below). As mentioned (see Section IV.c, p.33), it is recommended that numeric criteria under such conditions should be designed to protect 95% of the desirable metric, following U.S. EPA guidance (Stephan et al. 1985).

An illustration of a numeric TP criterion that could be derived from line-drawing is taken from Figure 8, wherein a chla inflection point of 21 µg/L between "good-acceptable" and "somewhat impaired" conditions for recreation in a group of Wisconsin lakes was indicated

from user perceptions. The average spring TP level associated with summer blooms averaging chla concentrations of ~20 µg/L was used to set as TP criteria that protected recreation in Wisconsin lakes, largely based on this information. Specifically, numeric TP criteria were set in various types of Wisconsin lakes so as to ensure that there would be only a 5% chance of a nuisance algal bloom with moderate biomass (20 µg chla/L).

What if the major focus for management is fish, which may be several higher trophic levels apart from the primary producers that are key responders to nutrient enrichment at or near the base of the food web? Such metrics (fish species richness, fish production, etc.) generally are indirectly related to nutrient concentrations, rather than directly related as are algal and plant key responders. For example, the designated use of a group of waterbodies considered the most important by resource managers might be maintenance of robust game (sport) fish populations. These species are common apex predators in reservoir food webs (e.g., Jones and Hoyer 1982, Michaletz et al. 2012). Sportfish abundance is only indirectly related to water-column nutrient concentrations through food abundance, habitat, and other factors. Ney (1996) reported a generalized relationship between sport fish biomass and TP concentrations in reservoirs (Figure 24). This "hump-shaped" curve gradually increases to a maximal value for sport fish biomass and suggests that, while sport fish biomass is not directly related to TP concentrations, food quality and habitat for sport fish decline at higher TP levels due to factors more directly related to TP, such as noxious algal blooms. Following the U.S. EPA recommendation given by Stephan et al. (1985), a numeric TP criterion for these reservoirs could be set at 70 µg/L.

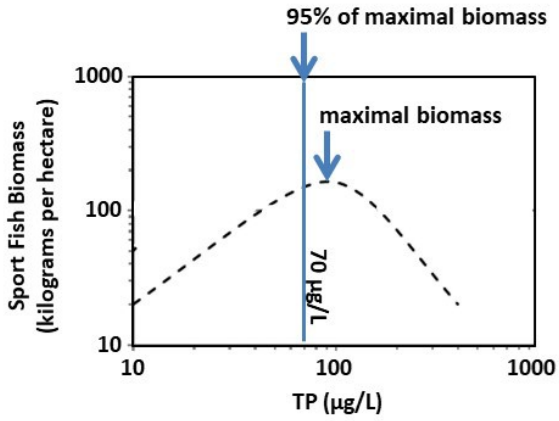


Figure 24. Generalized relationship sport fish biomass and water-column TP concentrations in some north temperate reservoirs. Adapted from Ney (1996).

In a third, somewhat different example, Blablil et al. (2017) developed a fish-based ecological index (comprised of nine metrics including one for cold-water trout species) to assess ecological quality of another set of reservoirs (Figure 25). Here, if resource managers targeted restoration or maintenance of "Good" ecological conditions, the water-column TP concentration in these reservoirs would need to be maintained at less than

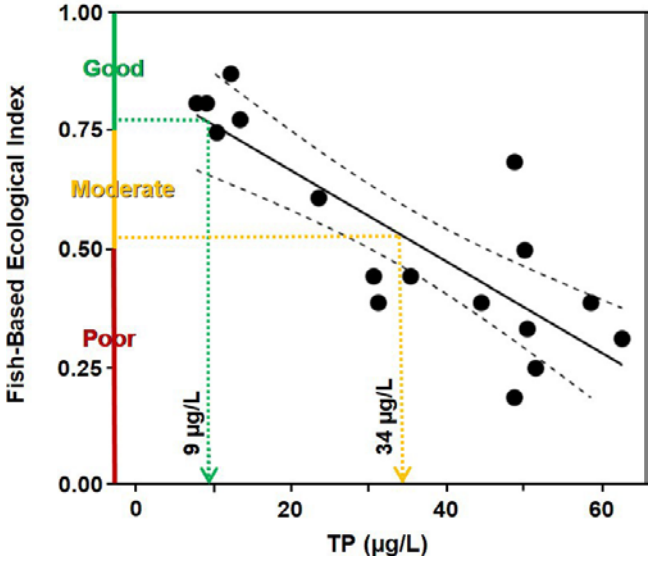


Figure 25. Relationship between a fish-based index of ecological quality and water-column TP concentration in another group of north temperate reservoirs. The dashed black lines are the 95% confidence intervals (linear regression: $R^2 = 0.68$, $p < 0.001$). Modified from Blabolil et al. (2017).

about 10 µg/L. Note that this approach includes a margin of safety, as the TP maximum correspond to a fish-based ecological index of ~0.77 rather than the lowest level for Good ecological conditions (0.75). Alternatively, if Moderate fish-based ecological conditions were considered by managers as sufficient to protect the designated use of these reservoirs for fish-based ecological health, then the TP level could be as high as about 35 µg/L, again including a margin of safety to ensure that the index remained within the Moderate range ($\geq$ 0.50).

V. Other Basic Guidance

V.a. The criteria should be developed from high-quality data, collected with sufficient frequency in the appropriate period: The most basic requirement for developing protective numeric criteria is a high-quality dataset for TP and TN concentrations. As explained above, high-quality dataset is defined as consisting of data less than a decade old, collected by a state-certified or otherwise reputable laboratory, and taken with sufficient frequency (at least monthly) to enable accurate assessment of the targeted value − for example, the 25th percentile of data for all streams in the area, or the 75th percentile of data for true reference or minimally impacted stream conditions (see below). A minimum frequency of monthly data collection is recommended to enable accurate assessment of general conditions (e.g., Robertson and Roerish 1999, Stansfield 2001). Chla data should have been collected consistently at about the same time of day, to reduce the variation caused by differing light/temperature regimes in the morning versus the afternoon. Dissolved oxygen data should have been collected with appropriately calibrated equipment; caution is warranted because if the instrument is not calibrated frequently enough, a layer of bacteria and other microbes colonize the equipment and compromise the data, and the data cannot be reliably "back-calculated."

Accurate assessment of minimally impacted conditions should be limited to the period of interest for which the numeric criteria are to be developed, typically the "summer" season or the "growing season" when noxious algal blooms mostly occur in response to nutrient. For example, May through October may be the growing season for certain crops, but noxious cyanobacteria blooms generally occur during the warmest months (June through August), and they do not occur or very rarely occur during October in north temperate waters of the upper Midwest. The "algal bloom season" should be restricted to the months when the blooms actually occur. In addition, assessment of average concentration within the period of interest should include baseflow, stormflow, and wetflow conditions (Hollabaugh and Harris 2004). Quarterly data (which could amount to only one sample per year when the period of most interest was a three-month summer season) are widely regarded as insufficient to assess mean nutrient concentrations in a particular surface water because they miss many storm vents and associated higher nutrient concentrations that would be detected with more frequent sampling (e.g., see Harmeson and Barcelona 1981, Stansfield 2001, Hollabaugh and Harris 2004).

The U.S. EPA's (2000a,b) nutrient ecoregions concept is important regarding data needs because the ecoregions cross state lines (Figure 11). If a state has only sparse or dated information on waterbodies of interest, data should be used from other states for waters within the same ecoregion. For example, if streams in central Iowa were targeted for numeric nutrient criteria but data for true minimally impacted conditions were not available, then

minimally impacted conditions for streams within the same ecoregion in Minnesota, Missouri, or Indiana should be used to develop the numeric criteria. Use of numeric criteria developed in nearby Wisconsin can be applied to similar types of waterbodies (e.g., streams) with similar land use/soils in Illinois, where agriculture or urbanization (Chicago influence) have contaminated waters to such an extent that very few minimally impacted streams remain (TetraTech, Inc. 2008).

V.b. Reference or minimally impacted conditions should be used in Approach #1: If true reference or minimally impacted conditions are used, the resulting criteria should protect the waterbody(s) from further degradation due to nutrient pollution (but see Section VI.b below).

V.c. Effects-based approaches should rely on protective statistical procedures for determining thresholds: Changepoint analysis has sometimes been shown to be inferior to AQRS to identify protective thresholds for sub-acute biological responses to stressors (Wang et al. 2007, Brenden et al. 2008; and see Figure 22 above). Changepoint analysis typically yields higher threshold nutrient concentrations for adverse effects on sensitive, beneficial biota than AQRS. Serial regression analysis can yield much higher nutrient thresholds than either changepoint analysis or AQRS because serial regression is confounded if the error terms are correlated. For example, in attempting a serial regression of BOD → chla → TP (considered in developing the Minnesota eutrophication criteria - see Heiskary et al. 2013), the error terms for chla and BOD would be correlated since chla contributes to BOD.

V.d. Numeric criteria should be developed for TN and TP, rather than only for response variable(s) such as chla: Focus only on a response variable such as chla for numeric criteria is reactive because there is commonly a lag period between TP and TN loads and supplies versus maximum phytoplankton response (e.g., Jones and Bachmann 1976, Withers et al. 2014, Bertani et al. 2016). Increased nutrient loads may not translate into increased algal biomass and other symptoms of nutrient pollution during the same season or the same year, depending on weather, flushing rates, light, and other environmental conditions. When these conditions become conducive for noxious algal outbreaks ("blooms"), then the higher nutrient supplies fuel the algal growth.

V.e. The dominant key responder(s) in the waterbody(s) of interest should be the targeted response variable(s) if included in numeric criteria: Suspended microalgal biomass as chla concentration is an appropriate response variable for use in many lakes, reservoirs, and larger rivers because the suspended microalgae or phytoplankton generally are the key responders in the open water (Wetzel 2001). These types of systems generally have a phytoplankton-driven food web (Indiana Department of Environmental Management undated, Zhen-Gang 2017). In contrast, smaller, fast-flowing streams generally are dominated by benthic algae (periphyton), and have a periphyton-driven food web (Indiana Department of Environmental Management undated, Zhen-Gang 2017). The benthic algae of smaller streams are the major responders to nutrient pollution, and they can provide a much better indication of stream health than the relatively sparse suspended microalgae (Welch et al. 1988, 1989). Suspended microalgae tend to be sparse in smaller streams because they are continually, rapidly flushed downstream before they can even reproduce. They are not an actual phytoplankton community; instead, they consist of algal cells that the turbulence has dislodged and suspended, and they came from the attached microalgae (Whitton 1975, Greeson 1982, Allan and Castillo 2007).

In some lakes, reservoirs, and quiet side-waters of rivers, the dominant key responders may be floating filamentous macroalgae, or noxious aquatic weeds (macrophytes). Numeric TN and TP criteria for such waters should be developed in such systems to reduce or minimize noxious growth of these primary producers. Depending on the waterbody, two key responder groups may need to be considered as the major response variables, such as both suspended microalgae and benthic filamentous algae in shallow waterbodies. Some lakes, for example, have developed noxious phytoplankton blooms and/or noxious benthic filamentous algal blooms in response to nutrient pollution, depending on the year and the climatic conditions (Lapointe et al. 2018).

V.f. Arithmetic means should be used to set numeric nutrient criteria, rather than geometric means or medians - Geometric means of a small number of samples do not reliably estimate average water quality conditions. A small number of samples (five or fewer) results in very large uncertainty in attempting to estimate the true mean (Sokal and Rolff 2012). Both the geometric mean and the median are less sensitive to outliers than the arithmetic mean (Sokel and Rolff 2012). As Parkhurst (1998) wrote,

> Concentrations of chemical substances and microorganisms are often averaged for scientific and regulatory purposes. Geometric means are sometimes used for these purposes, but they are biased low and do not represent components of mass balances properly. They should be abandoned in favor of arithmetic means, unless they are clearly shown to be preferable for specific applications. Arithmetic means are unbiased, easier to calculate and understand, scientifically more meaningful (at least for concentration data), and more protective of public health.

Geometric means are commonly used for assessment of fecal bacteria conditions because densities of fecal bacteria can vary by 100,000-fold, and geometric means depress the extreme variance. Nutrient concentrations vary by much less. The bias inherent in geometric means, that is, consistent underestimate of the true mean, is a serious problem in efforts to protect water quality. The geometric mean is always smaller than the arithmetic mean unless all numbers in the dataset are identical; then the geometric mean and the mean will be equal. Parkhurst (1998, p.98A) noted that "If geometric means are used with small datasets [$N \leq 5$], they should always [emphasis added] be bias-corrected."

V.g. Compliance should be based on independent consideration of TN, TP, and response variables: Silva (2010) wrote, "It is U.S. EPA's current position that exceedance of a causal water quality criterion, such as N or P, would require a determination of non-attainment regardless of whether the water is meeting a different water quality criterion, such as chlorophyll-a or other biological response criteria." Thus, TP, TN, chla, and other numeric criteria should each "stand alone" to protect waterbodies from nutrient pollution-related degradation. That is, if any one of the causal/response variables is exceeded, then the waterbody should be characterized as in violation of the numeric criteria.

Independent consideration of multiple criteria is necessary because nutrient pollution can detrimentally affect beneficial aquatic life through means other than increased algal growth. Eutrophication is defined as an increase in the rate of organic production in an aquatic ecosystem (see Burkholder and Glibert 2013, and references therein). As explained in Section

III.b.1, increased organic production in response to nutrient over-enrichment can be contributed by various primary producer groups, such as certain aquatic vascular plants, floating macroalgae, or benthic algae, rather than the suspended microalgae that are the only organisms included in the chla indicator.

V.h. Sampling locations should be selected carefully for assessment of compliance: Smaller streams typically are sampled in the channel midpoint where streambank influences are minimal but also where flow and potential dilution are greatest. Larger rivers require more thought in determining the best locations to sample, because noxious algal blooms develop and accumulate along the sides ("side-channel") rather than in the middle of the channel. For characterizing the water quality of larger rivers, side-channel data should be taken for nutrients as well as for response variables such as chla, since side-channel locations are where noxious blooms usually develop. The locations where measurements are to be taken in a lake or reservoir also should be specified. Otherwise, a waterbody could mistakenly receive site-specific or modified criteria based on measurements in the most nutrient-degraded area as "average" conditions applied to the entire waterbody.

VI. Doing It Wrong: Non-protective Actions in Developing Numeric Nutrient Criteria

Some common non-science-based actions that have been used to design weaker numeric nutrient criteria than warranted are summarized below. These actions should be avoided:

VI.a. Setting numeric criteria reactively, that is, when pollution-tolerant species predominate and pollution-sensitive species are rare or gone (i.e., eliminated from the system) - Several examples given above in Section IV-3.b.2 reflected this problem. Sophisticated statistical techniques such as AQRS should not be applied in an attempt to determine "protective thresholds" in waters that have been polluted because the analysis will only reveal threshold nutrient concentrations that, finally, adversely impact even the remaining tolerant species. Analyses should be scrutinized to ensure that threshold analysis emphasizes sensitive taxa rather than pollution-tolerant taxa. The latter approach would actually eliminate sensitive biotas that are valued ecosystem components (also see VI.d below).

VI.b. Failure to use minimally impacted conditions - The meaning of "minimally impacted" may seem obvious, but this important concept sometimes has been irrationally defined. For example, a state that lies partly within a heavily agricultural region formally defined minimally impacted conditions as streams that were not affected by point sources of nutrient pollution (McCollor and Heiskary 1993). Major nonpoint pollution from agriculture was not considered(!), so that streams which drained mostly agricultural watersheds but did not have impacts from sewage treatment plants were wrongly evaluated as "minimally impacted." Actual minimally impacted streams, without influence from both point and nonpoint sources, had much lower TP and TN concentrations than the agricultural nonpoint pollution-impaired streams that were formally defined as minimally impacted. Numeric criteria were developed based, in part, on highly polluted conditions as minimally impacted. Use of compromised minimally impacted conditions will wrongly skew the analysis toward poorer water quality as "acceptable" in supporting pollution-tolerant biota, while also not protecting the biota of higher-quality waters in the area. In addition, threshold conditions

cannot be accurately identified because the higher-quality waters will mostly be missing from comparison.

As a related error, in some cases numeric nutrient criteria have been developed for waterbodies within highly urbanized or dense agricultural areas, using the 25th percentile of all-streams data. The U.S. EPA (2000a-c) cautioned that in such highly compromised watersheds, a much lower percentile of data should be used in order to develop protective criteria, such as the 5th or 15th percentile of all-streams data.

VI.c. Rejection of data from true minimally impacted or reference waters in the same nutrient ecoregion of an adjacent state to set numeric nutrient criteria, when such waters are no longer available in the state of interest - As clearly shown in Figures 11, 13, and 14, nutrient ecoregions and sub-ecoregions do not "stop" at state lines. Protective numeric criteria can be set for surface waters in highly compromised areas (e.g., characterized by dense agriculture or high urbanization) by basing the criteria on minimally impacted or reference waters available in a nearby state. A state which lacks reference waters, yet relies only on data from that state, would be rejecting valid data and actually using invalid data – from waters already polluted well beyond minimally impacted conditions – to set standards.

VI.d. Use of a "novel" (made up!) procedure that defeats the purpose of the accepted statistical technique upon which the concocted procedure was based - Analyses should be examined to ensure that the statistical procedure being used, such as AQRS, is actually followed and the identified thresholds are used, rather than "suddenly adding" a new, non-protective step(s).

For example, the same state that mistakenly considered streams draining high-density agricultural watersheds as "minimally impacted" (Section VI-b above) also used AQRS in a hybrid approach to set numeric nutrient criteria (Heiskary et al. 2013). AQRS required selection of lambda (λ) values, which determine the amount of smoothing applied to the line drawn through the wedge-shaped biological data. The λ values were selected/adjusted based on how well the line (model) appeared to fit the outside curve and identified breakpoints. A statistical test (F test) was then used to assess goodness of fit of the data to the model. If the data was statistically significant at the first (most protective) breakpoint - that is, if the data fit the line well in that area - then the first breakpoint was used as the threshold TP concentration for the biological metric being analyzed. But if not, the much higher TP concentration halfway between the first and second actual thresholds indicated by the AQRS was used as the "threshold concentration" for setting the numeric criterion.

This procedure, newly conceived by the agency, was given the statistically authentic-sounding name, "midpoint interpolation" and was added at the end of the AQRS analysis. The name helped to camouflage the fact that clear, first breakpoints protective of sensitive taxa were ignored in favor of much less protective, higher TP concentration as an artificial "threshold" that was <u>not</u> indicated by the AQRS technique. Instead, AQRS indicated that the TP concentration used by the agency was at a much higher TP concentration along the TP gradient where the biological metric was rapidly declining. Here is why: As is the present reality for many U.S. surface waters, few waterbodies in the study had high taxa richness inclusive of sensitive taxa. Most sensitive taxa had been extirpated. Thus, there were few

waterbodies that included sensitive taxa. The regression line tended to fit most of the data, which were from the most waterbodies which no longer had sensitive species, and only roughly fit the sparse datapoints for the more pristine waters. Rather than using the first, protective threshold revealed by AQRS, the agency "slid halfway down the slope" between the first and second TP breakpoints and used that concentration as an artificial "threshold" (which was not an actual threshold) for the response of that biological metric to increasing TP concentration.

Figure 26 shows the same data as in Figure 22, but with added "midpoint interpolation" (indicated by Xs). For taxa richness, the "artificial threshold" TP concentration from "midpoint interpolation" was the same as the concentration indicated using changepoint analysis, 127 µg/L - but the first threshold concentration actually indicated by AQRS was ~110 µg/L. The alteration of AQRS by addition of "midpoint interpolation" was more serious for the second biological metric shown in Figure 26. AQRS again indicated a protective first threshold concentration of ~110 µg/L; changepoint analysis indicated a higher breakpoint at 182 µg TP/L; and the artificial "midpoint threshold" was more than 230 µg TP/L.

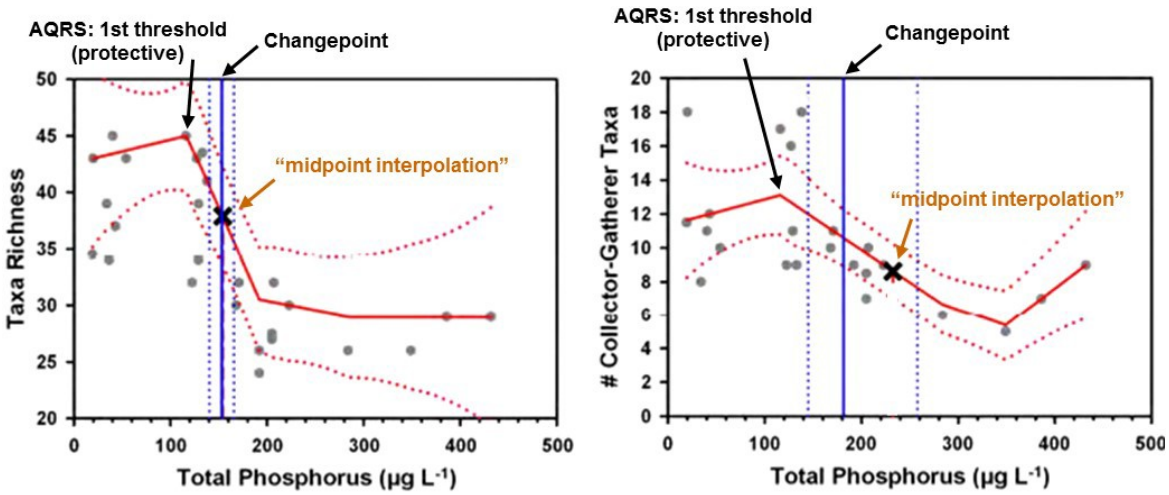


Figure 26. The same data shown in Figure 22, but with added "midpoint interpolations;" the artificial midpoint "thresholds" are indicated by an X. In both panels, the first protective threshold revealed by AQRS is at ~110 µg TP/L, and the artificial midpoint "threshold" is at a substantially higher TP level. Left panel: the artificial midpoint "threshold" is at the same TP concentration as the breakpoint from change-point analysis. Right panel: the artificial midpoint "threshold" is more than twice as high as the first protective TP threshold indicated by AQRS, and also substantially higher than the breakpoint from changepoint analysis. Thus, use of "midpoint interpolation" would eliminate protection of the higher-quality biological metrics characteristic of minimally impacted waters. Data from Heiskary et al. (2013; also see Heiskary and Bouchard 2015).

The agency maintained that it had used an advanced hybrid approach (minimally impacted conditions, changepoint analysis, and AQRS) which strongly supported the numeric criteria that were selected. Note, however, that at least two of the three analyses - use of waters that had been seriously compromised by agricultural pollution as "minimally impacted," and ignoring the first TP thresholds revealed by AQRS in favor of much higher midpoint artificial "thresholds" - had the overall effect of selecting for non-protective numeric TP criteria.

VI.e. Use of a response variable such as phytoplankton or benthic algal chlorophyll a as a "surrogate" for numeric TN and TP criteria - Some states have mistakenly described chlorophyll a as a surrogate for numeric TN and TP criteria, but chlorophyll a is not a reliable surrogate: TN and TP are strongly related to some water quality conditions that are not related, or only weakly related, to chlorophyll a, such as disinfectant byproduct (DBP) precursors (Veum 2006).

Chla has also been mistakenly described as effectively "integrating" the impacts of TN and TP [Missouri Department of Natural Resources 2016]. Other factors additionally influence algal biomass, such as temperature, light, and water motion (Whitton 1975, Wetzel 2001). Thus, chla is not a reliable integrator of TN and TP alone because other factors can confound the relationship between TN or TP and chla, so that algal response is temporarily depressed. That is why it is important to develop numeric criteria for the causal variables, TN and TP, themselves.

VI.f. Constructing a model to predict numeric criteria for nutrients and response variable(s), based on little or no actual data from the waterbody(s) of interest - There is an old saying in the modeling arena, "A model is only as good as the data used to build it." Yet, it has been common for models used to develop numeric eutrophication criteria to be based on few or no data for key parameters. The resulting model is unreliable at predicting nutrient or response variables.

A recent example is the modified, site-specific DO criterion for the Escatawpa River in Mississippi (Burkholder 2011, 2013). The sophisticated Environmental Fluid Dynamics Code/ Water Quality Analysis Simulation Program (EFDC/WASP) model was selected for use. The modelers unfortunately ignored the fact that the EFDC/WASP model requires substantial data from the system being modeled in order to make reliable predictions. Data were not available from the Escatawpa for nearly all of the ~28 parameters needed, so guesstimates for them were "plugged in" from other waterbodies, sometimes from waters with very different pollution histories in different regions, and with little or no justification. Values used for a key parameter that greatly influenced the model predictions, sediment oxygen demand (SOD), were selected based on sparse, inadequate samples taken during a season when SOD would have been low rather than elevated (Hatcher 1986). In the calibration/validation process, model predictions should be compared to actual, measured water quality data, but few or no data were available for the needed parameters – even key parameters such as DO. As a result, the model failed at reliably predicting DO concentrations. Analogously, if this was a game of darts, the model usually failed to place the darts on the board (Burkholder 2011).

Modeling requires specialized expertise, but interpreting model quality requires only sound, science-based logic. The key to model assessment lies in determining the underlying assumptions used to build the model, the validity of those assumptions, and the quality of the data used to construct it. The information provided about a model often includes a list of key assumptions, but it becomes the task of the reviewer to assess whether some key assumption(s) was not mentioned. Commonly, a major assumption omitted from the list is that extremely sparse data, or no data, from the system being modeled was "adequate," or that parameters such as algal productivity, taken from the mid-ocean (which unfortunately been done), were "sufficiently accurate" to construct a model of a turbid river.

VI.g. Use of erroneous assumptions and other non-science-based rationale - Various modeling approaches that have been used to develop numeric criteria for nutrients and/or response variables have not been science-based. Again, the quality of the data and the underlying assumptions can easily reveal whether the model is valid. In the effort to alter the DO criterion for the Escatawpa River in the above example, there was no certainty that the species considered in the model even occurred in the river; one low-oxygen-tolerant species was used to "represent" all of the fish species; growth of a coldwater invertebrate that does not occur in the entire warm temperate region which includes the Escatawpa was used to "represent" growth of all invertebrates in the river; and summer-spawning species were "forced" in the model to spawn during colder seasons when oxygen is plentiful, which allowed a false prediction of negligible impact from the modified, much lower DO criterion on aquatic life in the river (Burkholder 2011, 2013). The analysis violated U.S. EPA guidelines for data requirements (Stephan et al. 1985), and also violated well-established science principles. Such models, using inadequate data and non-science-based approaches, should be avoided in developing numeric criteria.

In that model as well, average "natural surface background" DO conditions were wrongly decreased by including low DO data from depths greater than five feet, which was supposed to have been the lowest depth considered for surface conditions according to the state regulations. The lower DO from deeper waters was then wrongly modeled as the "surface" condition. More than 80% of the actual data exceeded 3 mg/L but, as a result of these and other "sleight of hand" actions described above, the model predicted that the "natural" condition of the Escatawpa River was 3 mg/L. The criterion that was eventually adopted, 3.7 mg/L (average daily, May through October), was modified from 5 mg/L and much lower than it should have been to protect the designated use of the Escatawpa River for aquatic life.

VII. Summary

This white paper explains the scientific background of nutrient pollution impacts on freshwaters, and the pressing need for numeric TP and TN criteria to protect their designated uses. Five approaches suggested by the U.S. EPA for use in developing numeric nutrient criteria are described and illustrated using examples. These approaches fall within two major lines of reasoning: The first line of reasoning seeks to set criteria based on what was the natural, pre-development or minimally impacted levels of nutrients. The second line of reasoning sets criteria based on studies of how nutrient pollution harms the designated uses of waterbodies (e.g., aquatic life support, fishing, swimming, drinking source-water, etc.).

The approach selected largely must depend on the data available for the waterbody(s) of interest (Dodds and Welch 2000). The dataset size/frequency recommended to develop numeric nutrient criteria for each of the various approaches should be compared to the data in hand for the system(s) to assess whether the data are sufficient for use with an approach being considered. If sufficient water quality (TN, TP, chlorophyll a) data are available for waterbody(s) of interest (e.g., data collected at least monthly during the summer growing season, over the past 5-10 years, for nutrients and chlorophyll), then numeric nutrient criteria should be developed based on localized conditions following accepted science-based procedures (Dodds and Welch 2000). This basically describes U.S. EPA's reference condition

(minimally impacted) approach, using nutrient sub-ecoregions rather than the much larger nutrient ecoregions, or using a river basin as in Suplee et al. (2015).

If a strong dataset is available for sensitive macroinvertebrate and fish biota, then stressor-response relationships and thresholds should be developed as well, using AQRS or other suitable statistical technique supported by peer-reviewed literature. It is recommended that if possible, multiple approaches should be applied as an internal check (e.g., Wagenhoff et al. 2012; and see Appendix 1). Where clear thresholds for identifying protective criteria are not evident, and/or the metric changes gradually as nutrient pollution increases, then selection of protective numeric nutrient criteria requires a line-drawing procedure informed by the desired conditions for the metric, as adopted by the U.S. EPA in setting other protective numeric criteria (Stephan et al. 1985).

The numeric criteria indicated by each of the five described approaches likely will vary depending on the approach and the biological metric(s) used. Considering the uncertainty inherent in understanding the complex relationships between nutrients and various response variables, the lower, more protective nutrient criteria indicated by the two analyses should be used (e.g., Dodds et al. 2010).

References Cited

Alexander, R.B., R.A. Smith, G.E. Schwarz, E.W. Boyer, J.V. Nolan, and J.W. Brakebill (2008) Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River basin. Environmental Science and Technology 42: 822-830.

Allan, J.D. and M.M. Castillo (2007) Stream Ecology: Structure and Function of Running Waters, 2nd edition. Springer, Dordrecht, Germany.

Anderson, M.J. (2008) Animal-sediment relationships revisited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. Journal of Experimental Marine Biology and Ecology 366: 16-27.

Apostel, A.M. (2016) Re-evaluating the Development of Phosphorus Loading Restrictions: Maumee River Case Study. M.S. thesis, The Ohio State University, Columbus. Available at: https://etd.ohiolink.edu/!etd.send_file?accession=osu1469108284&disposition=inline.

Baker, M.E. and R.S. King (2010) A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution 1: 25-37.

Barnwell, T.O., L.C. Brown, and R.C. Whittemore, 2004. Importance of field data in stream water quality modeling using QUAL2E-UNCAS. Journal of Environmental Engineering 130: 643-647.

Basu. B.K. and F.R. Pick (1996) Factors regulating phytoplankton and zooplankton biomass in temperate rivers. Limnology and Oceanography 41: 1572-1577.

Benke, A.C. and C.E. Cushing (2005) Rivers of North America. Elsevier Academic Press, Burlington (MA) (p.864).

Bertani, I., D.R. Obenour, C.E. Steger, C.A. Stow, A.D. Gronewold, and D. Scavia (2016) Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. Journal of Great Lakes Research 42: 1184-1192.

Bierman, V.J., J.V. DePinto, D.W. Dilks, P. Moscus, T.A.D. Slawecki, C.F. Bell, S.C. Chapra, and K.F. Flynn (2013) Modeling Guidance for Developing Site-Specific Nutrient Goals. LINK1T11. Water Environment Research Foundation, Alexandria (VA).

Biggs, B.J.F. (2000) Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. Journal of the North American Benthological Society 19: 17-31.

Biggs, B.J.F. and R.A. Smith (2002) Taxonomic richness of stream benthic algae: effects of flood disturbance and nutrients. Limnology and Oceanography 47: 1175-1186.

Blabolil, P., M. Říha, D. Ricard, J. Peterka, M. Prchalová, M. Vašek, M. Čech, J. Frouzová, T. Jůza, M. Muška, M. Tušer, V. Draštík, Z. Sajdlová, M. Šmejkal, L. Vejřík, J. Matěna, D.S. Boukal, D. Ritterbusch, and J. Kubecka (2017). A simple fish-based approach to assess the ecological quality of freshwater reservoirs in Central Europe. Knowledge and Management of Aquatic Ecosystems 418: 53 (12 pp.).

Black, R., P. Moran, and J. Frankforter (2011) Response of algal metrics to nutrients and physical factors and identification of nutrient thresholds in agricultural streams. Environmental Monitoring and Assessment 175: 397-417.

Boyer, G.L. (2007) The occurrence of cyanobacterial toxins in New York lakes: Lessons from the MERHAB-Lower Great Lakes Program. Lake and Reservoir Management 23: 153-160.

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

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

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

Bukhari, H., J. Elkin, M. Hemken, N. Lichten, R. Pettit, and S. Shattuck (2015) Evaluation of Different Approaches for Controlling Phosphorus Pollution in the Maumee River Watershed. Masters Project, School of Natural Resources and Environment, University of Michigan, Ann Arbor. Available at: https://deepblue.lib.umich.edu/handle/2027.42/111006.

Burkholder, J.M. (1992) Phytoplankton and episodic suspended sediment loading: Phosphate partitioning and mechanisms for survival. Limnology and Oceanography 37: 974-988.

Burkholder, J.M. (2002) Cyanobacteria, pp. 952-982. In: Encyclopedia of Environmental Microbiology, by G. Bitton (ed.). Wiley Publishers, New York.

Burkholder, J.M. (2009) Harmful algal blooms, pp. 264-285. In: Encyclopedia of Inland Waters, Volume 1, by G.E. Likens (ed.) Elsevier, Oxford, United Kingdom.

Burkholder, J.M. (2010) Analysis of the draft site-specific water quality standard for Lake Byllesby. Prepared for the Minnesota Center for Environmental Advocacy, St. Paul.

Burkholder, J.M. (2011) Affidavit of JoAnn M. Burkholder regarding a revised dissolved oxygen criterion for the lower Escatawpa River, Mississippi. Submitted to the Mississippi Department of Environmental Quality, Jackson.

Burkholder, J.M. (2013) Supplemental Affidavit of JoAnn M. Burkholder regarding a revised dissolved oxygen criterion for the lower Escatawpa River, Mississippi. Submitted to the Mississippi Department of Environmental Quality, Jackson.

Burkholder, J.M. (2014) Evaluation of Phosphorus Pollution from MWRD's Three Largest Water Reclamation Plants to Receiving Waters, from the Chicago Area Water System to the Peoria Pool. Report for Plaintiffs National Resources Defense Council, Inc., Sierra Club, Inc. and Prairie Rivers Network - NRDC, et al. v. MWRD, No. 11-2937.

Burkholder, J.M. and P.M. Glibert (2013) Eutrophication and oligotrophication. In: Encyclopedia of Biodiversity, Vol. 2, second edition, by Simon Levin (ed.). Academic Press, New York.

Burkholder JM, Gobler C, O'Neill J (2018, in press) Cyanobacteria. In: Harmful Algal Blooms and Their Management: A Compendium Desk Reference, by S.E. Shumway, J.M. Burkholder, and S.L. Morton (eds.). Wiley, New York.

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

Caley, M.J., M.H. Carr, M.A. Hixon, T.P. Hughes, G.P. Jones, and B.A. Menge (1996) Recruitment and the local dynamics of open marine populations. Annual Review of Ecology and Systematics 27: 477-500.

Camargo, J.A. and A. Alonso (2006) Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment. Environment International 32: 831-849.

Camargo, J.A., A. Alonso, and A. Salamanca (2005) Nitrate toxicity to aquatic animals: A review with new data for freshwater invertebrates. Chemosphere 58: 1255-1267.

Canton, S., C. Wolf, and N. Burbank (2012) Derivation of nutrient criteria using existing community thresholds. Presentation, Annual Meeting of the Society for Freshwater Science, Louisville (KY).

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

Carlson, R.E. and J. Simpson (1996) A Coordinator's Guide to Volunteer Lake Monitoring Methods. North American Lake Management Society, Madison.

Carleton, J.N., R.A. Park and J.S. Clough (2009) Ecosystem modeling applied to nutrient criteria development in rivers. Environmental Management 44: 485-492.

Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications 8: 559-568.

Caskey, B.J., A. Bunch, M. Shoda, and R.J. Miltner (2012) Identifying Nutrient Reference Sites in Nutrient-Enriched Areas: Using Algal, Invertebrate, and Fish-Community Measures to Identify Stressor Thresholds in Indiana Rivers and Streams, 2005-9. U.S. Geological Survey Scientific Investigations Report 2011-5009. USGS, Reston (VA).

Chambers, P.A., D.J. McGoldrick, R.B. Brua, C. Vis, J.M. Culp, and G.A. Benoy (2012) Development of environmental thresholds for nitrogen and phosphorus in streams. Journal of Environmental Quality 41: 7-20.

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

Chaudhuri, P. and J.S. Marron (1999) SiZer for exploration of structures in curves. Journal of the American Statistical Association 94: 807-823.

Chen, J. and A.K. Gupta (2012) Parametric Statistical Change Point Analysis, 2nd edition. Birkhäuser Basel – Springer Science+Business Media, LLC, Heidelberg, Germany.

Chorus, I., and J. Bartram (eds.) (1999) Toxic Cyanobacteria in Water – A Guide to Their Public Health Consequences, Monitoring and Management. E & FN Spon for the World Health Organization, New York.

Clements, W.H., N.K.M. Vieira, and D.L. Sonderegger (2010) Use of ecological thresholds to assess recovery in lotic ecosystems. Journal of the North American Benthological Society 29: 1017-1023.

Cloern, J.E. (2001) Our evolving conceptual model of the coastal eutrophication problem. Marine Ecology Progress Series 210: 223-253.

Cooke, G.D., E.B. Welch, S.A. Peterson, and S.A. Nichols (2005) Management and Restoration of Lakes and Reservoirs, 3rd edition. CRC Press, Boca Raton (FL).

Cormier, S.M., J.F. Paul, R.L. Spehar, P. Shaw-Allen, W.J. Berry, and G.W. Suter (2008) Using field data and weight of evidence to develop water quality criteria. Integrated Environmental Assessment and Management 4: 490-504.

Correll, D.L. (1998) The role of phosphorus in the eutrophication of receiving waters: A review. Journal of Environmental Quality 27: 261-266.

Craft, C. 2013. Emergent macrophyte biomass production, pp. 137-153. In: Methods in Biogeochemistry of Wetlands, by R.D. DeLaune, K.R. Reddy, C.J. Richardson, and J.P.

Megonigal (eds.) Soil Science Society of America (SSSA) Book Ser. 10. SSSA, Madison.

Daily, J.P. N.P. Hitt, D.R. Smith, and C.D. Snyder (2012) Experimental and environmental factors affect spurious detection of ecological thresholds. Ecology 93: 17-23.

Déath, G. (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83: 1105-1117.

Déath, G. and K.E. Fabricius (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81: 3178-3192.

Denton, G.M., D.H. Arnwine, and S.H. Wang (2001) Development of Regionally-Based Interpretations of Tennessee's Narrative Nutrient Criterion. A criterion development document prepared for the Tennessee Water Quality Control Board. Division of Water Pollution Control, Tennessee Department of Environment and Conservation, Nashville.

Dillon, P.J. and F.H. Rigler (1974) The phosphorus-chlorophyll relationship in lakes. Limnology and Oceanography 19: 767-773.

Dillon, P.J. and F.H. Rigler (1975) A test of a simple phosphorus nutrient budget model predicting the phosphorus concentration in lake water. Journal of the Fisheries Research Board of Canada 31: 1771-1778.

Dodds, W.K. (2006) Eutrophication and trophic state in rivers and streams. Limnology and Oceanography 51: 671-680.

Dodds, W.K. (2007) Trophic state, eutrophication and nutrient criteria in streams. TREE (Trends in Ecology and Evolution) 22: 669-676.

Dodds, W.K., W.H. Clements, K. Gido, R.H. Hilderbrand, and R.S. King (2010) Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. Journal of the North American Benthological Society 29: 988-997.

Dodds, W.K., J.R. Jones, and E.B. Welch (1998) Suggested classification of stream trophic state: distributions of temperate stream types by chlorophyll, total nitrogen, and phosphorus. Water Research 32:1455-1462.

Dodds, W.K. and V.H. Smith (2016) Nitrogen, phosphorus, and eutrophication in streams. Inland Waters 6: 155-164.

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 Science 59: 865-874. and

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 Science 63: 1190-1191.

Dodds, W.K., V.H. Smith, and B. Zander (1997) Developing nutrient targets to control benthic

chlorophyll levels in streams: a case study of the Clark Fork River. Water Research 31: 1738-1750.

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

Edsall, D.A. and C.E. Smith (1991) Effect of oxygen supersaturation on rainbow trout fed with demand feeders. The Progressive Fish-Culturist 53: 95-97.

Evans-White, M.A., W.K. Dodds, D.G. Huggins, and D.S. Baker (2009) Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams. Journal of the North American Benthological Society 28: 855-868.

Evans-White, M.A., B.E. Haggard, and J.T. Scott (2014) A review of stream nutrient criteria development in the United States. Journal of Environmental Quality 42: 1002-1014.

Flynn, K.F., M.W. Suplee, S.C. Chapra, and H. Tao (2015) Model-based nitrogen and phosphorus (nutrient) criteria for large temperate rivers: 1. Model development and application. Journal of the American Water Resources Association 51: 421-446.

Force, G. and E. Macbeth (2002) Volunteer Water Quality Monitoring of Lake Pepin and Spring Lake: a Citizen Monitoring Network on the Mississippi River, 1994-1998 Five-Year Summary. Minnesota Boundary Area Commission report to Metropolitan Council Environmental Services.

Frey, J.W., A.H. Bell, J.A. Hambrook Berkman, and D.L. Lorenz (2011) Assessment of Nutrient Enrichment by Use of Algal-, Invertebrate-, and Fish-Community Attributes in Wadeable Streams in Ecoregions surrounding the Great Lakes. USGS Scientific Investigations Report 2011-5009. National Water-Quality Assessment Program, USGS, Reston (VA).

Garvey, J.E., E.A. Marschall , and R.A. Wright (1998) From star charts to stoneflies: detecting relationships in continuous bivariate data. Ecology 79:442-447.

Geider, R.J., and J. La Roche (2002) Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. European Journal of Phycology 37: 1-17.

Glass, P.W. (2006) Development of Use-Based Chlorophyll Criteria for Recreational Uses of Reservoirs. Alan Plummer Associates, Inc., Austin.

Glibert, P.M. and J.M. Burkholder (2018, in press) Causes of blooms. In: Harmful Algal Blooms and Their Management: A Compendium Desk Reference, by S.E. Shumway, J.M. Burkholder, and S.L. Morton (eds.). Wiley, New York.

Glibert, P.M., D. Fullerton, J.M. Burkholder, J.C. Cornwell, and T.M. Kana (2011) Ecological stoichiometry, biogeochemical cycling, invasive species and aquatic food webs: San Francisco Estuary and comparative systems. Reviews in Fisheries Science 19: 358-417.

Gobler, C.J., J.M. Burkholder, T.W. Davis, M.J. Harke, T. Johengen, C. Stow, and D.B. Van de

Waal (2016) The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae 54: 87-97.

Graham, L.E., J.M. Graham, and L.W. Wilcox (2010) Algae, 2nd edition. Pearson Benjamin Cummings, San Francisco.

Greeson, P.E. 1982. Biota and Biological Principles of the Aquatic Environment. Geological Survey Circular 848-A. USGS, U.S. Department of the Interior, Washington, DC.

Hansen, E., M. Christ, M. Janes, J. Hankins, and N. Gillies (April 21, 2006) Recommended Nutrient Criteria for West Virginia Lakes. Submitted to the Division of Water and Waste Management, West Virginia Department of Environmental Protection, Charleston.

Harmeson and Barcelona (1981) Sampling Frequency for Water Quality Monitoring. State Water Survey (SWS) Contract Report 279. Illinois Department of Energy and Natural Resources, SWS Division, Aquatic Chemistry Section, Champaign.

Hatcher, K.J. (1986) Sediment Oxygen Demand: Processes, Modeling, and Measurement. Institute of Natural Resources, University of Georgia, Athens.

Hecky, R.E., P. Campbell, and L.L. Hendzel (1993) The stoichiometry of carbon, nitrogen, and phosphorus in particulate matter of lakes and oceans. Limnology and Oceanography 38: 709-724.

Hecky, R.E. and P. Kilham (1988) Nutrient limitation of phytoplankton in freshwater and marine environments: A review of recent evidence on the effects of enrichment. Limnology and Oceanography 33: 796-822.

Hedman, C.J., W.R. Krick, and D.A. Karner Perkins (2008) New measurements of cyanobacterial toxins in natural waters using high performance liquid chromatography coupled to tandem mass spectrometry. Journal of Environmental Quality 37: 1817-1824.

Heiskary, S.A. and R.W. Bouchard, Jr. (2015) Development of eutrophication criteria for Minnesota streams and rivers using multiple lines of evidence. Freshwater Science 34: 574-592.

Heiskary, S., R.W. Bouchard Jr., and H. Markus (2013) Minnesota Nutrient Criteria Development for Rivers - Update of November 2010 Report. Minnesota Pollution Control Agency, St. Paul.

Heiskary, S. and H. Markus (2003) Establishing Relationships Among In-Stream Nutrient Concentrations, Phytoplankton Abundance and Composition, Fish IBI and Biochemical Oxygen Demand in Minnesota USA Rivers. Environmental Analysis & Outcomes Division, Minnesota Pollution Control Agency, St. Paul.

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

Hendry, G.A.F., J.D. Houghton, and S.B. Brown (1987) The degradation of chlorophyll - a biological enigma. Tansley Review #11. New Phytologist 107: 255-302.

Henze, M., P. Harremoes, J. Jansen, and E. Arvin (1997) Wastewater Treatment - Biological and Chemical Processes, 2nd edition. Springer, New York.

Hilderbrand, R.H., R.M. Utz, S.A. Stranko, and R.L. Raesly (2010) Applying thresholds to forecast potential biodiversity loss from human development. Journal of the North American Benthological Society 29: 1009-1016.

Hillebrand, H., G. Steinert, M. Boersma, A. Malzahn, C.L. Meunier, C. Plum, and R. Ptacnik (2013) Goldman revisited: faster-growing phytoplankton has lower N:P and lower stoichiometric flexibility. Limnology and Oceanography 58: 2076–2088.

Ho, J.C. and A.M. Michalak (2017) Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. Journal of Great Lakes Research 43: 221-228.

Hollabaugh, C.L. and R.R. Harris (2004) Refinement of sampling methods for analysis of nutrients in surface waters. Environmental Geosciences 11: 112-128.

Houser, J.N., D.W. Bierman, R.M. Burdis, and L.A. Soeken-Gittinger (2010) Longitudinal trends and discontinuities in nutrients, chlorophyll, and suspended solids in the Upper Mississippi River: implications for transport, processing, and export by large rivers. Hydrobiologia 651: 127-144.

Houser, J.N. and W. Richardson (2010) Nitrogen and phosphorus in the Upper Mississippi River: Transport, processing, and effects on the river ecosystem. Hydrobiologia 640: 71-88.

Hoyer, M.V., C.D. Brown, and D.E. Canfield Jr. (2004) Relations between water chemistry and water quality as defined by lake users in Florida. Lake and Reservoir Management 20: 240-248.

Hoyer, M.V. and D.E. Canfield Jr. (1996) Largemouth bass abundance and aquatic vegetation in Florida lakes: An empirical analysis. Journal of Aquatic Plant Management 34: 23-32.

Huang, X., K-f. Li, J. Du, and R. Li (2010) Effects of gas supersaturation on lethality and avoidance responses in juvenile rock carp (Procypris rabaudi) Tchang). Journal of Zhejiang University - Science B (Biomedicine & Biotechnology) 11: 806-811.

Huggins, D.G. and J. Anderson (2005) Dissolved Oxygen Fluctuation Regimes in Streams of the Western Corn Belt Plains Ecoregion. Report No. 130. Kansas Biological Survey, University of Kansas, Lawrence.

Huston, M.A. (1994) Biological Diversity: The Coexistence of Species on Changing Landscapes. Cambridge University Press, Cambridge, United Kingdom.

Huston, M.A. (2004) Management strategies for plant invasions: manipulating productivity, disturbance, and competition. Diversity and Distributions 10: 167-178.

Indiana Department of Environmental Management. Undated. Algae in Aquatic Ecosystems. Fact sheet, at: https://www.in.gov/idem/files/factsheet_owq_sw_algae_aquatic.pdf.

Jeppesen, E., M. Søndergaard, M. Meerhoff, T.L. Lauridsen, and J.P. Jensen (2005) Shallow lake restoration by nutrient loading reduction - some recent findings and challenges ahead. Hydrobiologia 584: 239-252.

Jones, J.R. and R.W. Bachmann (1976) Prediction of phosphorus and chlorophyll levels in lakes. Journal of the Water Pollution Control Federation 48: 2176-2184.

Jones, J.R. and M. Hoyer (1982) Sportfish harvest predicted by summer chlorophyll-a concentration in Midwestern lakes and reservoirs. Transactions of the American Fisheries Society 111: 176-179.

Jones, J.R., M.F. Knowlton, D.V. Obrecht, A.P. Thorpe, and J.D. Harlan (2009) Role of contemporary and historic vegetation on nutrients in Missouri reservoirs: implications for developing nutrient criteria. Lake and Reservoir Management 25: 111-118.

Kalff, J. (2002) Limnology. Prentice-Hall, Upper Saddle River (NJ).

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

Kirk, J.T.O. (1994) Light and Photosynthesis in Aquatic Ecosystems, 2nd edition. Cambridge University Press, Cambridge, United Kingdom.

Koenker, R. (2005) Quantile Regression. Cambridge University Press, Cambridge, United Kingdom.

Koenker, R. (2016) Quantreg: Quantile Regression, 2016. R package version 5.21. Available at: https://CRAN.R-project.org/package=quantreg.

Kratzer, C.R. and P.L. Brezonik (1981) A Carlson-type trophic state index for nitrogen in Florida lakes. Water Resources Bulletin 17: 713-715.

Kutner, M.H., C.J. Nachtsheim, and J. Neter (2004) Applied Linear Regression Models. McGraw-Hill Irwin, New York.

Lake Erie Improvement Association (2012) Strategic Plan for Lake Erie Partners: Sustaining Healthy Waters for Lake Erie's Economy. Lake Erie Improvement Association in consultation with Hull & Associates, Inc., Oregon (OH).

Lapointe, B.E., J.M. Burkholder, and K. Van Alstyne (2018) Harmful macroalgal blooms in a changing world: Causes, impacts, and management. In: Harmful Algal Blooms and Their Management: A Compendium Desk Reference, by S.E. Shumway, J.M. Burkholder, and S.L. Morton (eds.). Wiley, New York.

Leland, H.V. and S.D. Porter (2000) Distribution of benthic algae in the upper Illinois River basin in relation to geology and land use. Freshwater Biology 44: 279-301.

Liebig, J. (1847) Chemistry in its Application to Agriculture and Physiology, 4th edition. Taylor &

Walton Publishers, London, United Kingdom.

Lindon, M.J. and S.A. Heiskary (2008) Blue-Green Algal Toxin (Microcystin) Levels in Minnesota Lakes. Environmental Bulletin No. 11. Minnesota Pollution Control Agency, St. Paul.

Lindon, M.J. and S.A. Heiskary (2009) Blue-green algal toxin (microcystin) levels in Minnesota lakes. Lake and Reservoir Management 25: 240-252.

Mackay, R. (2006) River and stream ecosystems of Canada, chapter 3. In: River and Stream Ecosystems of the World, by C.E. Cushing, K.W. Cummins, and G.W. Minshall (eds.). University of California Press, Berkeley.

Mallin, M.A., V.L. Johnson, S.H. Ensign, and T.A. MacPherson (2006) Factors contributing to hypoxia in rivers, lakes, and streams. Limnology and Oceanography 51: 690-701.

Manner, R. (2014) 2014 Illinois Water. Presentation. Urbana & Champaign Sanitary District and Illinois Association of Wastewater Agencies - Nutrient Committee.

Mazzotti, F.J. and J.J. Vinci (2010) Validation, verification, and calibration: using standardized terminology when describing ecological models. University of Florida / Institute of Food and Agricultural Sciences. Available at: http://edis.ifas.ufl.edu/pdffiles/UW/UW25600.pdf.

McCollor, S. and S. Heiskary (1993) Selected Water Quality Characteristics of Minimally Impacted Streams from Minnesota's Seven Ecoregions. Minnesota Pollution Control Agency, St. Paul.

McConkey, S., A. Bartosova, L.-S. Lin, K. Andrew, M. Machesky, and C. Jennings (2004) Fox River Watershed Investigation - Stratton Dam to the Illinois River: Water Quality Issues and Data Report to the Fox River Study Group, Inc. (Chapter 7). Prepared by the Illinois State Water Survey, Champaign.

Meals, D. W., S. A. Dressing, and T. E. Davenport (2010) Lag time in water quality response to best management practices: A review. Journal of Environmental Quality 39: 85-96.

Metcalf & Eddy, Inc. (1991) Wastewater Engineering, Treatment, Disposal, and Reuse, 3rd edition. McGraw-Hill, Inc., New York.

Michael, H.J., K.J. Boyle, and R. Bouchard. 2000. Does the measurement of environmental quality affect implicit prices estimated from hedonic models. Land Economics 76: 283-298.

Michaletz, P.H., D.V. Obrecht, and J.R. Jones (2012) Influence of environmental variables and species interactions on sport fish communities in small Missouri impoundments. North American Journal of Fisheries Management 32: 1146- 1159.

Mills, W.B., G.L. Bowie, T.M. Grieb, K.M. Johnson, and R. Whittemore (1986) Handbook: Stream Sampling for Waste Load Allocation Applications. Report #EPA/625/6-86/013. Office of Research and Development, U.S. EPA, Washington, DC.

Miltner, R.J. (2010) A method and rationale for deriving nutrient criteria for small rivers and streams in Ohio. Environmental Management - DO 10.1007/s00267-010-9439-9.

Minihan, K.(2017) Water Quality Standards – What's Coming Up for Lakes? Presentation at the Wisconsin Lakes Convention (April 7). Wisconsin Department of Natural Resources, Madison.

Minnesota Pollution Control Agency (MPCA) (2010) In the Matter of the Proposal to Develop a Site Specific Standard for the Byllesby Reservoir – Dakota & Goodhue Counties, Minnesota. Findings of Fact – Conclusions of Law and Order. MPCA, St. Paul, 12 pp.

Mississippi Department of Environmental Quality (MDEQ) (2009) Total Maximum Daily Load - Total Nitrogen and Total Phosphorus for the Pearl River - Pearl River Basin. MDEQ, Jackson.

Mississippi River Basin Watershed Nutrient Task Force (2010) Hypoxia 101. U.S. EPA, Washington, DC. Available at: http://water.epa.gov/type/watersheds/named/msbasin/hypoxia101.cfm.

Missouri Department of Natural Resources (MoDNR) (2016) Rationale for Missouri Numeric Nutrient Criteria for Lakes. MoDNR, Jefferson City.

Morgan, A.M., T.V. Royer, M.B. David, and L.E. Gentry (2006) Relationships among nutrients, chlorophyll-a, and dissolved oxygen in agricultural streams in Illinois. Journal of Environmental Quality 35: 1110-1117.

Morgan, R.P. and K.M. Kline, 2011. Nutrient concentrations in Maryland non-tidal streams. Environmental Monitoring and Assessment 178: 221-225.

Muradian, R. (2001) Ecological thresholds: a survey. Ecological Economics 38: 7-24.

National Research Council (2000) Clean Coastal Waters – Understanding and Reducing the Effects of Nutrient Pollution. National Academy Press, Washington, DC.

Nebeker, A.V., F.D. Baker, and S.L. Weitz (1981) Survival and adult emergence of aquatic insects in air-supersaturated water. Journal of Freshwater Ecology 1: 243-250.

Ney, J.J. (1996) Oligotrophication and its discontents: effects of reduced nutrient loading on reservoir fisheries. American Fisheries Society Symposium 16: 285-295.

Nordin R.N. (1985) Water quality criteria for nutrients and algae (technical appendix). Water Quality Unit, Resource Quality Section, Water Management Branch, B. C. Ministry of Environment, Victoria, British Columbia, Canada.

Parker, T.M. (2013) Effects of the Interaction of Environmental Factors (Hypoxia and Ammonia) on Fish. M.S. thesis, Ohio State University, Columbus.

Parkhurst, D.F. (1998, February 1) Arithmetic versus geometric means for environmental concentration data. Environmental Science & Technology / News, pp. 92 A - 98 A.

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: 1211-1222.

Perrin, C.J., N.T. Johnston and S.C. Samis (1988) Effects of treated sewage effluent on periphyton and zoobenthos in the Cowichan River, British Columbia. Canadian Technical Report of Fisheries and Aquatic Science 1591, 64 pp.

Pitcairn, C.E.R. and H.A. Hawkes (1973) The role of phosphorus in the growth of Cladophora. Water Research 7: 159-171.

Prairie, Y.T., C.M. Duarte, and J. Kalff (1989) Unifying nutrient-chlorophyll relationships in lakes. Canadian Journal of Fisheries and Aquatic Science 46: 1176-1182.

Qian, S.S. and T.F. Cuffney (2012) To threshold or not to threshold? That's the question. Ecological Indicators 15: 1-9.

Qian, S.S., R.S. King, and C.J. Richardson (2003) Two statistical methods for the detection of environmental thresholds. Ecological Modelling 166: 87-97.

R Development Core Team (2009) R: A language and environment for statistical computing. Vienna, Austria.

Raschke, R.L. and D.A. Schultz (1987) The Use of the Algal Growth Potential Test for Data Assessment. Journal of the Water Pollution Control Federation 59: 222-227.

Redfield, A.C. (1958) The biological control of chemical factors in the environment. American Scientist 46: 205-222.

Robertson, D.M., D.J. Graczyk, P.J. Garrison, L. Wang, G. LaLiberte, and R. Bannerman (2006) Nutrient Concentrations and Their Relations to the Biotic Integrity of Wadeable Streams in Wisconsin. United States Geological Survey (USGS) Professional Paper 1722. USGS, Reston (VA).

Robertson, D., and E. Roerish (1999) Influence of various water quality sampling strategies on load estimates for small streams. Water Resources Research 35: 3747-3759.

Robertson, D.M., G.E. Schwarz, D.A. Saad, and R.B. Alexander (2009) Incorporating uncertainty into the ranking of Sparrow Model nutrient yields from Mississippi/ Atchafalaya River basin watersheds. Journal of the American Water Resources Association 45: 534-549.

Robertson, D.M., Weigel, B.M., and D.J. Graczyk (2008) Nutrient Concentrations and Their Relations to the Biotic Integrity of Nonwadeable Rivers in Wisconsin. 2008. United States Geological Survey (USGS) Professional Paper 1754. USGS, Reston (VA).

Rook, S. (2012, Dec. 31) Testing on Minnesota River shows positive results. St. Peter Herald. Available at: http://www.pca.state.mn.us/index.php/about-mpca/mpca-news/featured-stories/water-quality-improves-in-minnesota-river.html.

Royer, T.V., M.B. David, L.E. Gentry, C.A. Mitchell, K.M. Starks, T. Heatherly II, and M.R.

Whiles (2008) Assessment of chlorophyll-a as a criterion for establishing nutrient standards in the streams and rivers of Illinois. Journal of Environmental Quality 37: 437-447.

Rykiel, E.J. Jr. (1996) Testing ecological models: the meaning of validation. Ecological Modelling 90: 229-244.

Sakamoto, M. (1966) Primary production by the phytoplankton community in some Japanese lakes and its dependence on lake depth. Archiv fur Hydrobiologie 62: 1-28.

Schindler, D.W., S.R. Carpenter, S.C. Chapra, R.E. Hecky, and D.M. Orihel (2016) Reducing phosphorus to curb lake eutrophication is a success. Environmental Science & Technology 50: 8923-8929.

Sheeder, S.A. and B.M. Evans (2004) Estimating nutrient and sediment threshold criteria for biological impairment in Pennsylvania watersheds. Journal of the American Water Resources Association 40: 881-888.

Silva, S.J. (2010) Letter to Mr. Thomas Danielson, Maine Department of Environmental Protection, March 15. Water Quality Section Branch, U.S. EPA Region 1, Boston.

Smeltzer, E. and S. Heiskary (1990) Analysis and application of lake user survey data. Lake and Reservoir Management 6: 109-118.

Smith, A.J., R.L. Thomas, J.K. Nolan, D.J. Velinsky, S. Klein, and B.T. Duffy (2013) Regional nutrient thresholds in wadeable streams of New York State protective of aquatic life. Ecological Indicators 29: 455-467.

Smith, A. J., and C. Tran (2010) A weight-of-evidence approach to define nutrient criteria protective of aquatic life in large rivers. Journal of the North American Benthological Society 29: 875-891.

Smith, C.S. and J.W. Barko (1990) Ecology of Eurasian watermilfoil. Aquatic Plant Management 28: 55-64.

Smith, D.G., G.F. Croker, and K. McFarlane (1995) Human perception of water appearance: clarity and colour for bathing and aesthetics. New Zealand Journal of Marine and Freshwater Research 29: 29-43.

Smith, V.H. (1982) The nitrogen and phosphorus dependence of algal biomass in lakes: an empirical and theoretical analysis. Limnology and Oceanography 27: 1101-1112.

Smucker, N.J., M. Becker, N.E. Detenbeck, and A.C. Morrison (2013) Using algal metrics and biomass to evaluate multiple ways of defining concentration-based nutrient criteria in streams and their ecological relevance. Ecological Indicators 32: 51-61.

Sokal, R.R. and F.J. Rolff. 2012. Biometry, 4th edition. W. H. Freeman, New York.

Sonderegger, D.L., H. Wang, W.H. Clements, and B.R. Noon (2009) Using SiZer to detect thresholds in ecological data. Frontiers in Ecology and the Environment 7: 190-195.

Stancheva, R. and R.G. Sheath (2016) Benthic soft-bodied algae as bioindicators of stream water quality. Knowledge and Management of Aquatic Ecosystems 417, 15. DOI: 10.1051/kmae/2016002

Stanley, E.H and J.T. Maxted (2008) Changes in the dissolved nitrogen pool across land cover gradients in Wisconsin streams. Ecological Applications 18: 1579-1590.

Stansfield, B. (2001) Effects of sampling frequency and laboratory detection limits on the determination of time series water quality trends. New Zealand Journal of Marine and Freshwater Research 35: 1071-1075.

Stephan, C.E., D.I. Mount, D.J. Hansen, J. Gentile, G.A. Chapman, and W.A. Brungs (1985) Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. U.S. EPA, Duluth (MN).

Stephenson, K. (2007) Recreational user perceptions of lake/reservoir water quality, Section II-E. In: Nutrients in Lakes and Reservoirs – A Literature Review for Use in Nutrient Criteria Development, by J.L. Walker, T. Younos, and C.E. Zipper. VWRRC Special Report SR34-2007. Virginia Water Resources Research Center, Virginia Tech, Blacksburg.

Stevenson, R.J., B.J. Bennett, D.N. Jordan, and R.D. French (2012) Phosphorus regulates stream injury by filamentous green algae, DO, and pH with thresholds in responses. Hydrobiologia 695: 25-42.

Stevenson, R.J., B.H. Hill, A.T. Herlihy, L.L. Yuan, and S.B. Norton (2008) Algae-P relationships, thresholds, and frequency distributions guide nutrient criterion development. Journal of the North American Benthological Society 27: 783-799.

Stevenson, R.J., S.T. Rier, C.M. Riseng, R.E. Schultz, and M.J. Wiley (2006) Comparing effects of nutrients on algal biomass in streams in two regions with different disturbance regimes and with applications for developing nutrient criteria. Hydrobiologia 561: 149-165.

Stoner, N. 2011 (March 1). Letter to R.F. Poltak, New England Water Pollution Control Commission, available at https://www.epa.gov/sites/production/files/2015-01/documents/newengland-interstate-nutrients-letter.pdf.

Stumm, W. and J.J. Morgan (1995) Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters, 3rd edition. John Wiley & Sons, New York.

Suplee, Michael W., Kyle F. Flynn, and Steven C. Chapra (2015) Model-based nitrogen and phosphorus (nutrient) criteria for large temperate rivers: 2. Criteria derivation. Journal of the American Water Resources Association 51: 447-470.

Suplee, M.W., V. Watson, M. Teply, and H. McKee (2009) How green is too green? Public opinion of what constitutes undesirable alga[e]l levels in streams. Journal of the American Water Works Association 45: 123-140.

Tafangenyasha, C., B.E. Marshall, and L.T. Dube (2010) The diurnal variation of the physicochemical parameters of a lowland river flow in a semi-arid landscape with human

interferences in Zimbabwe. International Journal of Water Resources and Environmental Engineering 2: 148-163.

Tetra Tech, Inc. (2008) Analysis of Illinois Stream and River Nutrient and Biological Data for the Nutrient Scientific Technical Exchange Partnership Support (N-STEPS). Prepared for the Office of Science and Technology of U.S. EPA, Washington, DC. Tetra-Tech, Inc., Owings Mills (MD).

Toms, J.D. and M.L. Lesperance (2003) Piecewise regression: A tool for identifying ecological thresholds. Ecology 84: 2034-2041.

United States Environmental Protection Agency (U.S. EPA) (1998) National Strategy for the Development of Regional Nutrient Criteria. Report #EPA 822-R-98-002. U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000a) Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs. Report EPA-822-B00-001. Office of Water and Office of Science and Technology, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000b) Nutrient Criteria Technical Guidance Manual: Rivers and Streams. Report EPA-822-B00-002. Office of Water and Office of Science and Technology, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000c) Nutrient Criteria Technical Guidance Manual. Report #EPA-822-B-00-002. U.S. EPA, Washington, DC. .

United States Environmental Protection Agency (U.S. EPA) (2000d) Ambient Water Quality Criteria Recommendations - Information Supporting the Development of State and Tribal Nutrient Criteria - Rivers and Streams in Nutrient Ecoregion IX: Southeastern Temperate Forested Plains and Hills. Report # EPA 822-B-00-019. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000e) Ambient Water Quality Criteria Recommendations - Information Supporting the Development of State and Tribal Nutrient Criteria - Rivers and Streams in Nutrient Ecoregion VI: Corn Belt and Northern Great Plains. Report #EPA 822-B-00-017. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000f) Ambient Water Quality Criteria Recommendations - Information Supporting the Development of State and Tribal Nutrient Criteria - Lakes and Reservoirs in Nutrient Ecoregion VIII: Nutrient Poor Largely Glaciated Upper Midwest and Northeast. Report #EPA 822-B-00-010. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000g) Ambient Water Quality Criteria Recommendations - Information Supporting the Development of State and Tribal Nutrient Criteria - Rivers and Streams in Nutrient Ecoregion IX - Southeastern Temperate Forested Plains and Hills. Office of Water and Office of Science and Technology, US EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2000h) Ambient Water Quality Criteria Recommendations - Information Supporting the Development of State and Tribal Nutrient Criteria - Rivers and Streams in Nutrient Ecoregion VII – Mostly Glaciated Dairy Region. Office of Water and Office of Science and Technology, US EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2001a) Memorandum: Development and Adoption of Nutrient Criteria into Water Quality Standards, by G. Grubbs. WQSP-01-01. Office of Science and Technology, Washington, DC.

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

United States Environmental Protection Agency (U.S. EPA) (2009a) Guidance on the Development, Evaluation, and Application of Environmental Models. Report #EPA/100/K-09/003. Office of the Science Advisor, Council for Regulatory Environmental Modeling, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2009b, January 14) Letter to M. Sole, Secretary, Florida Department of Environmental Protection, from B.H. Grumbles, Assistant Administrator, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2010) Using Stressor-response Relationships to Derive Numeric Nutrient Criteria. Report #EPA-820-S-10-001. Office of Science and Technology and Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2013) Guiding Principles on an Optional Approach for Developing and Implementing a Numeric Nutrient Criterion that Integrates Causal and Response Parameters. Report #EPA-820-F-13-039. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (U.S. EPA) (2015a) Preventing Eutrophication: Scientific Support for Dual Nutrient Criteria. Fact sheet #EPA-820-S-15-001. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (EPA). 2015b. Health Effects Support Document for the Cyanobacterial Toxin Microcystins. Report #EPA-820R15102. Health and Ecological Criteria Division, Office of Water, U.S. EPA, Washington, DC. Available at: https://www.epa.gov/sites/production/files/2015-06/documents/microcystins-support-report-2015.pdf.

United States Environmental Protection Agency (EPA). 2015c. Health Effects Support Document for the Cyanobacterial Toxin Cylindrospermopsin. Report #EPA-820R15103. Health and Ecological Criteria Division, Office of Water, U.S. EPA, Washington, DC. Available at: https://www.epa.gov/sites/production/files/2015-06/documents/cylindrospermopsin-support-report-2015.pdf.

United States Environmental Protection Agency (EPA). 2015d. Human Health Ambient Water

Quality Criteria: 2015 Update. Report #EPA 820-F-15-001.Office of Water and Office of Toxicology, U.S. EPA, Washington, DC. Available at: https://www.epa.gov/sites/production/files/2015-10/documents/human-health-2015-update-factsheet.pdf.

United States Environmental Protection Agency (U.S. EPA) (2016a) Renewed Call to Action to Reduce Nutrient Pollution and Support for Incremental Actions to Protect Water Quality and Public Health. Memorandum from J. Beauvais, Deputy Assistant Administrator, to State Environmental Commissioners and State Water Directors. Office of Water, U.S. EPA, Washington, DC.

United States Environmental Protection Agency (EPA). 2016b. Human Health Recreational Ambient Water Quality Criteria or Swimming Advisories for Microcystins and Cylindrospermopsin. Report #EPA 822-P-16-002. Office of Water, U.S. EPA, Washington, DC. Available at: https://www.epa.gov/wqc/draft-human-health-recreational-ambient-water-quality-criteria-andor-swimming-advisories.

United States Environmental Protection Agency (U.S. EPA) (2017) Nutrient Pollution - The Problem. Available at: https://www.epa.gov/nutrientpollution/problem.

Utz, R.M., R.H. Hilderbrand, and D.M. Boward (2009) Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. Ecological Indicators 9: 556-567.

Vallentyne, J.R. (1974) The Algal Bowl - Lakes and Man. Miscellaneous Special Publication 22. Dept. of the Environment, Fisheries and Marine Service, Ottawa, Canada.

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

Veum, K.S. (2006) Disinfection By-product Precursors and Formation Potentials of Missouri Reservoirs. A thesis presented to the Faculty of the Graduate School University of Missouri-Columbia.

Vollenweider, R.A. (1968) Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, with Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication. Technical Report DA 5/SCI/68.27. Organisation for Economic Cooperation and Development, Paris, France.

Wagenhoff, A., J. Clapcott, and E. Goodwin (2012) Review of River Ecosystem Threshold Models. Cawthron Institute, Nelson, New Zealand.

Walker, W.W. Jr. (1983) Significance of eutrophication in water supply reservoirs. Journal of the American Water Works Association 75: 38-42.

Wang, L., D.M. Robertson, and P.J. Garrison (2007) Linkages between nutrients and assemblagees of macroinvertebrates and fish in wadeable streams: Implication to nutrient criteria development. Environmental Management 39: 194-212.

Weitkamp, D. E., and M. Katz. 1980. A review of dissolved gas supersaturation literature. Transactions of the American Fisheries Society 109: 659-702.

Welch, E.B., R.R. Horner, and C.R. Patmont (1989) Prediction of nuisance periphytic biomass: a management approach. Water Research 23: 401-405.

Welch, E.B., J.M. Jacoby, R.R. Horner, and M.R. Seeley (1988) Nuisance levels of periphytic algae in streams. Hydrobiologia 157: 161-168.

Wetzel, R.G. (2001) Limnology - Lake and River Ecosystems, 3rd edition. Academic Press, New York.

Wetzel, R.G. and G.E. Likens (2000) Limnological Analyses, 3rd edition. Springer-Verlag, New York.

Weigel, B.M., and D.M. Robertson (2007) Identifying biotic integrity and water chemistry relations in nonwadeable rivers of Wisconsin: toward the development of nutrient criteria. Environmental Management 40: 691-708.

Whitton, B.A. (ed.) (1975) River Ecology. Studies in Ecology (Oxford, England), Volume 2. University of California Press, Berkeley.

Wilcock, R.J., J.W. Nagels, G.B. McBride, K.J. Collie, B.T. Wilson, and B.A. Huser (2010) Characterisation of lowland streams using a single-station diurnal curve analysis model with continuous monitoring data for dissolved oxygen and temperature. New Zealand Journal of Marine and Freshwater Research 32: 67-79.

Wisconsin Department of Natural Resources (WDNR) (2017) Guidance for Implementing Wisconsin's Phosphorus Water Quality Standards for Point Source Discharges, 2nd edition. Guidance Number 3400-2011-02. WDNR, Madison.

Withers, P.J.A., C. Neal, H.P. Jarvie, and D.G. Doody (2014) Agriculture and eutrophication: Where do we go from here? Sustainability 6: 5853-5875.

World Health Organization [WHO] (2003) Guidelines for Safe Recreational Waters, Volume 1 – Coastal and Fresh Waters, Chapter 8: Algae and cyanobacteria in fresh water. WHO, Geneva, Switzerland.

Xenopoulos, M.A., E. Porter-Goff, D.E. Spooner, C.J. Williams, and H.F. Wilson (2013) The prevalence of nonlinearity and detection of ecological thresholds for dissolved organic matter, nutrients and ecosystem functions across land use gradients. Proceedings of the Annual Meeting of the Association for the Sciences of Limnology and Oceanography (abstract).

Young, R.G., C.D. Matthaei, and C.R. Townsend (2008) Organic matter breakdown and ecosystem metabolism: functional indicators for assessing river ecosystem health. Journal of the North American Benthological Society 27: 605-625.

Zhen-Gang, J. (2017) Hydrodynamics and Water Quality: Modeling Rivers, Lakes, and Estuaries. Wiley, New York.

Appendix 1 – More about statistics used to assess stressor-response relationships

This paper has greatly simplified the approaches that have been used to test for non-linearity and thresholds in stressor-response relationships (as examples, see Dodds et al. 2010, Wagenhoff et al. 2012). Diverse methods are available for finding thresholds and characterizing nonlinearity between a driver (e.g., TP or TN) gradient and a variable describing the ecological state (e.g., sensitive macroinvertebrate species). Dodds et al. (2010) and Wagenhoff et al. (2012) reviewed some techniques (e.g, Table A1), several of which were included in Section IV. A brief description of several other statistical methods follows.

The various techniques described here and in Section IV are available in automated computer programs. Because these programs involve various user choices, the user should be experienced in statistical threshold analysis so that appropriate precautions can be taken and pitfalls can be avoided (e.g., Daily et al. 2012). For thresholds to be meaningful, statistical understanding is required both to apply an appropriate model and to meet the assumptions of the statistical test (Wagenhoff et al. 2012). As another caution, univariate ecological variables that are often used in monitoring, such as total macroinvertebrate taxa richness or EPT (ephemeropteran, plecopteran, and trichopteran macroinvertebrate taxa richness) can sometimes over-simplify complex systems with many species and can minimize or "hide" ecological thresholds that actually exist. Thus, use of multiple individual species responses may also be needed to detect ecological thresholds in some waterbodies (Wagenhoff et al. 2012).

The old adage, "Absence of evidence does not mean evidence of absence" also applies here: The absence of conspicuous breakpoints in stressor-response relationships should not be taken as "proof" of the absence of ecological thresholds in response to a nutrient stressor, because a focused "point" for ecological response is only one potential indicator of ecosystem state (Wagenhoff et al. 2012). Critical nutrient concentrations or thresholds of harm can also be based on concerns about community changes or loss of ecosystem functions (Cormier et al. 2008, Young et al. 2008), and can be quantified using a benchmark approach (Wagenhoff et al. 2012; below).

A. Definitions

Additive Quantile Regression Analysis with Smoothing (AQRS) – similar to linear QR, except that instead of fitting a single line to the data, AQRS fits a regression line to subsets of the data. AQRS can be used to identify thresholds in addition to fitting the outside of wedge-shaped data (below). AQRS is a powerful tool but requires substantial data to work well. Interpretations may still be limited since correlation does not necessarily mean causation; and if N and P concentrations are significantly correlated (which is common), their effects cannot be separated. Automated programs are available for data analysis using AQRS.

Benchmark approach - a non-statistical method for detecting an ecological threshold (Wagenhoff et al. 2012), which can be used to define thresholds of harm (Cormier et al. 2008). The threshold can be based on knowledge of a quantifiable cause-effect relationship between a nutrient stressor and a response variable with stakeholder value, such as the

abundance of a key species for recreational fishers. The benchmark approach sets a threshold at the stressor value where there is an expected deviation in an ecological response from a maximum expected or reference value.

Bivariate data - data for each of two variables, where each value of one variable is paired with a value of the other variable.

Breakpoint – the place (point) where one functional relationship between a driver and a response variable changes to another; often used interchangeably with threshold or changepoint (e.g., Daily et al. 2012).

Changepoint analysis – a statistical tool used to identify homogeneity within a dataset; or, the process of detecting distributional changes within observations. May be performed as a parametric analysis if assuming that the data follow a known distribution (Chen and Gupta 2012), or as a nonparametric analysis without that assumption; the latter is more typical of analyses applied to aquatic systems affected by nutrient pollution.

Derivative - of a function $y = f(x)$ of variable x is a measure of the rate at which the value y of the function changes as variable x changes.

Driver – an abiotic or biotic change over time or space that influences the ecosystem structure or function (Dodds et al. 2010).

Linear relationship – a linear relationship with direct proportionality such that, when plotted, the data indicate straight line. A given change in an independent variable will always produce a corresponding change in the dependent variable.

Locally weighted scatterplot smoothing (LOESS, LOWESS smooth, locally weighted smoothing or locally weighted scatterplot smoothing) – a nonparametric tool in regression analysis that creates a smooth line through a data plot (U.S. EPA 2010). It is usually used with messy (highly variable or "noisy") data or sparse data, or when weak relationships obscure any line of best fit. Smoothing techniques such as LOESS find a curve of best fit to the data without assuming a priori a distribution shape.

Nonlinear relationship – type of relationship between two variables wherein a change in one variable will not correspond with a constant change in the other variable. The graph of a nonlinear relationship will be a curve, or a set of curves, or a set of two or more straight lines, but it will not be one straight line.

Nonparametric – refers to statistical techniques that do not require the data to follow a normal or other distribution; nonparametric statistical tests do not assume any specific distribution. These tests can be applied when the major relationship between the stressor and the response is not linear.

Nonparametric Changepoint Analysis (nCPA) – changepoint analysis conducted without the assumption that the data follow a known distribution. Unlike another common, closely related threshold technique (two-dimensional Kolmogorov-Smirnov test, 2DKS; Garvey et

al. 1998), nCPA allows calculation of an uncertainty estimate around the threshold value (Evans-White et al. 2009).

Normal distribution – a relationship between two variables that can be described as a symmetrical bell-shaped curve, wherein most of the data occur around the middle of the variable range, with symmetrically tapering "tails" of lower and higher values to the left and right of the middle.

Parametric – assumes some type of distribution to the data, usually a normal distribution.

Polynomial regression – fits a nonlinear relationship between the value of x (dependent variable) and the corresponding conditional mean of y (independent variable; the conditional mean is the expected value of y). A polynomial trend line is a curved line.

Quantile Regression Analysis (QR) – Can be used to identify nutrient threshold concentrations that protect the designated uses of waterbodies. QR is appropriate for use with a dataset when the assumptions for SLR are not applicable (cannot be met).

Residual – the difference between the measured value of the dependent (response) variable and the predicted value in a regression model (see Figure A1). That is, residual = the observed value minus the predicted (estimated) value; the residuals are the vertical distances between the points of the dataset and the fitted line. Each data point has one residual. The residuals show whether a mathematical function, such as a line in a linear regression, accurately represents a dataset.

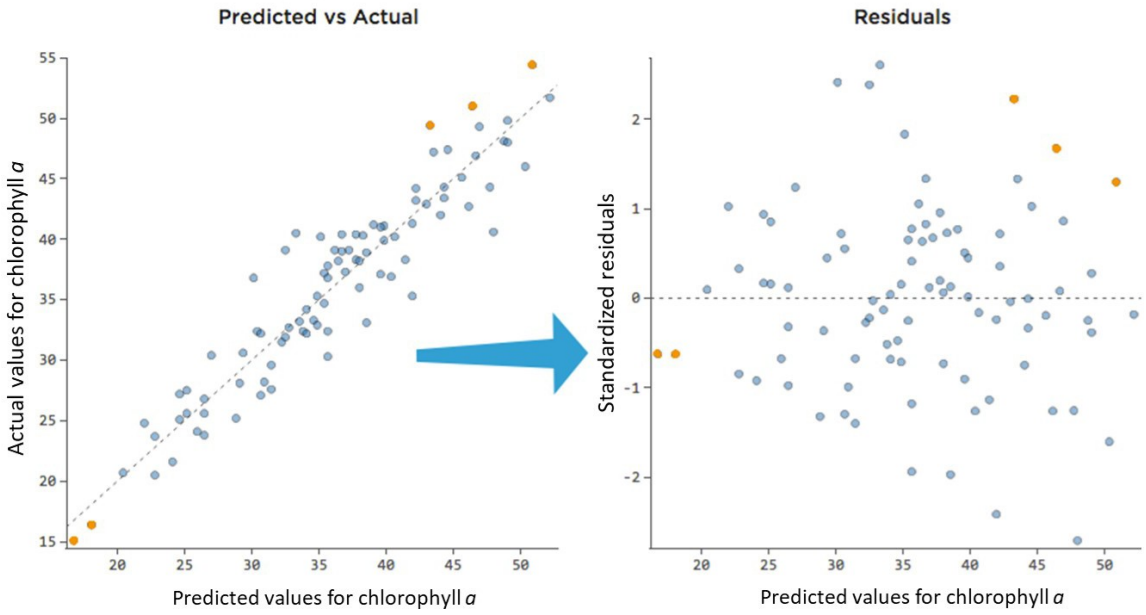


Figure A1. Example of a residual plot for chlorophyll a. The estimated or predicted values are the values that fall on the line. If the points are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data. Otherwise, a non-linear model should be used.

Simple linear regression (SLR) – a two-dimensional, linear regression model with one

independent variable (x, such as the causal variable TP) and one dependent variable (y, such as the response variable chlorophyll a). The linear function (non-vertical straight line) accurately predicts the dependent variable values as a function of the independent variable values. The adjective "simple" reflects the fact that the dependent (outcome) variable is related to a single predictor variable.

Standard deviation (σ or s) – a measure used to quantify the amount of variation or dispersion in a dataset. It is equal to the square root of the variance.

Stressor - an environmental driver that exceeds its normal range of variation due to human activity, and adversely affects ecosystem structure or function (Wagenhoff et al. 2012).

Stressor-response (driver-state) - the causal relationship between stressor (here, TN and/or TP) and response variables (U.S. EPA 2000a,b; Wagenhoff et al. 2012).

Threshold – The point at which there is an abrupt change in an ecological quality, property, or phenomenon, or where small changes in a stressor (driver) can produce large responses in the ecosystem (Groffman et al. 2006 in Dodds et al. 2010).

Wedge-shaped data (Figure A2) – refers to plots of field data for biota which show unequal variation at different nutrient levels. A biological attribute of interest (e.g., fish species richness) may be limited by nutrients on the outer or upper edge of the wedge, whereas inside the wedge, the biological attribute is limited by other environmental factor(s). This unequal variation of the biota response at different nutrient levels can make field data inappropriate for analysis by simple linear regression (SLR) or related techniques that assume equal variance across the data.

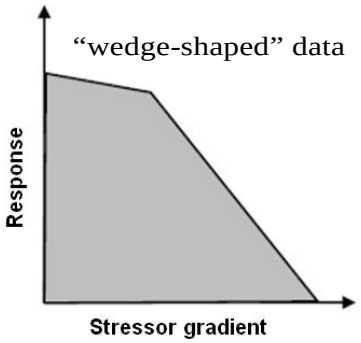


Figure A2. The general "wedge-shaped" outline of data for a given biological metric within the gray area

Variance – the average of the squared differences (considering each data point) from the mean. To calculate the variance, compute the difference between each data point and the mean; square each difference; add these values together; and then divide by the number of data points.

B. Statistical methods Used to Identify Nonlinear and Threshold Relationships

Breakpoint regression analysis (segmented or piecewise or "broken-stick" regression) is suitable for datasets which, when graphed after fitting a scatterplot smoother (e.g., a polynomial regression LOESS function) to the data to estimate the number of potential breakpoints (Wagenhoff et al. 2012) (Table A1). The data plot should reveal at least two clearly different linear relationships with a sudden sharp change in direction or slope at each. The boundaries between the segments are referred to as breakpoints. If a breakpoint is detected, the residuals from the fitted smooth function should be checked to verify that the assumption of constant error variance is met; if not, appropriate data transformation can sometimes be used to meet the assumption (Wagenhoff et al. 2012). Alternatively, data with

Table A1. Some statistical methods used to identify nonlinear and threshold relationships. Modified from Dodds et al. (2010).

Statistical Method	Description
Breakpoint [piecewise] regression	Statistically determines whether two relationships (usually linear) fit the data better than one; common in modern statistical packages.
Cumulative frequency distributions	Detect changes in expected distributions; can be combined with analytical methods to assess differences among distributions and acceptable methods of change (Paul and McDonald 2005, Utz et al. 2009, Hilderbrand et al. 2010).
Nonlinear curve fitting	Searches methods used to fit a user-defined equation; common in modern statistical packages.
Nonparametric changepoint analysis (nCPA)	A test for change in variance, closely related to 2-dimensional Kolmogorov-Smirnov test; adapted to calculate empirical confidence intervals (King and Richardson 2003, Qian et al. 2003). (Note: Threshold indicator taxa analysis [TITAN], used to indicate change in community structure across a gradient [Baker and King 2010], is a form of nCPA.)
Quantile regression	Characterization of changes in variance; includes quantile regression trees and quantile regression splines (Chaudhuri and Loh 2002, Cade and Noon 2003, Anderson 2008).
Recursive partitioning (regression tree analysis)	Predictors can be categorical or continuous; response can be univariate (classification and regression tree, CART) or multivariate (multivariate regression tree, MRT) (Breiman et al. 1984, Déath and Fabricius 2000, Déath 2002).
Significant zero crossings (SiZer)	Derivative-based nonparametric approach using a smoothing function (Sonderegger et al. 2009).
2-dimensional Kolmogorov-Smirnov test (2-DKS)	Nonparametric test for changes in variance; predicts a threshold in the driver and the response variable (Garvey et al. 1998).

non-normal distribution can be analyzed using generalized linear models if an appropriate error structure can be discerned (Toms and Lesperance 2003, Qian and Cuffney 2012). The statistical significance of the identified breakpoint can then be tested using various methods (see Wagenhoff et al. 2012).

Breakpoint analysis was used by Dodds et al. (2010) to identify a breakpoint in a relationship between TP and macroinvertebrate species richness in Central Plains streams (Figure A3). As another example, Black et al. (2011) identified breakpoints in responses of benthic algal and macroinvertebrate metrics along nutrient gradients, based on streams in the western U.S.

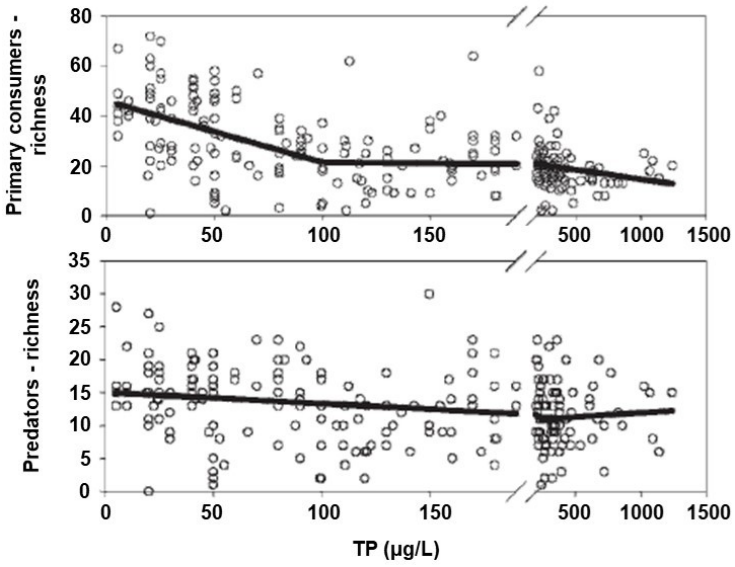


Figure A3. Example of breakpoint regression analysis of (upper graph) macroinvertebrate consumer species richness and (lower graph) predator species richness as a function of water-column TP concentration. This dataset is from 237 U.S. Central Plains stream sites (KS, NE, MO) in spring and autumn (Evans-White et al. 2009, in Dodds et al. 2010).

Cumulative frequency distributions are used to determine the number of observations that are above (or below) a particular value in a dataset. The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors, such as in the examples in Table A2 and Figure A4.

Table A2. Generic example of a cumulative frequency distribution (type = data grouping for this hypothetical example).

Type	Frequency	Cumulative Frequency
1-100	10	10
101-200	14	24
201-300	12	36
301-400	19	55

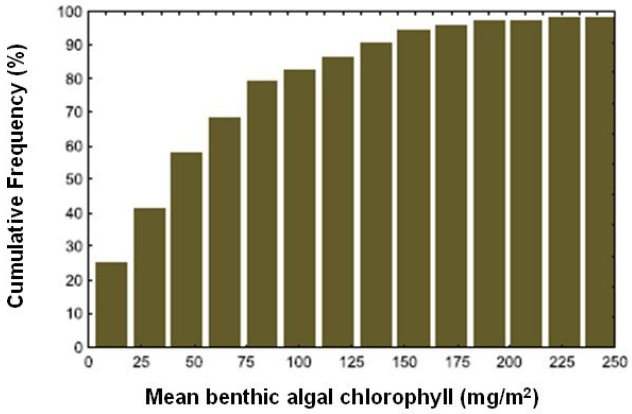


Figure A4. Cumulative frequency distribution for mean benthic algal chlorophyll from a literature dataset on temperate streams, compiled by Dodds et al. (2002).

CFD analysis requires a fairly large sample size and a broad range of stressor values (Wagenhoff et al. 2012). CFDs have been used to detect thresholds for responses of multiple taxa to stressors s (Utz et al. 2009). The technique compares observed versus expected CFD curves of negatively responding taxa across a stressor gradient to identify a lower limit for detection of an adverse effect (called the initiation-of-impact threshold) and an upper limit beyond which a taxon ceases to occur (extirpation threshold). The lower and upper limits were defined by Utz et al. (2009) at the points where the observed cumulative frequency of a single taxon deviates by $\geq$ 1% and $\geq$ 95%, respectively, from the expected cumulative frequency. Such limits can be modified as appropriate to suite the level of uncertainty (Wagenhoff et al. 2012).

As an illustration of the application of CFDs, the trophic classification scheme for streams and rivers that was developed by Dodds et al. (1998), based on suspended microalgal and benthic algal chlorophyll a (Table 2), was based on a CFD of the data. Three trophic state categories were developed based on the lower, middle, and upper thirds of the data distributions, and were assigned to oligotrophic, mesotrophic, and eutrophic categories, respectively. As a related example, the U.S. EPA (2000b) described how a CFD of sites by nutrient concentrations could be divided into three segments representing the stream quality categories of reference, acceptable (fair), and impaired. The U.S. EPA (2000b) recommended that a sample size of 30 waterbodies within a given class is needed for developing numeric nutrient criteria, as smaller sample sizes may be much more affected by extreme values in the dataset.

3b.2. Nonparametric Changepoint Analysis (nCPA, or nonparametric deviance reduction)

This technique, which is based on regression tree analysis (Breiman et al. 1984), can be used to estimate the position of a threshold (changepoint or breakpoint) in a univariate response variable affected by a stressor. The change in the response variable can be considered as a shift in the mean, variance, or rate of change. A form of nCPA known as the deviance reduction method is commonly used, which sorts observatons along a stressor gradient and identifies the response variable concentration where the deviance is minimized (Qian et al. 2003). Thus, when scatter plots visually suggest that a threshold or sudden change exists in the relationship between a stressor and a response, changepoint analysis can be used to identify the causal variable concentration (e.g., TN or TP) at which the change occurs (Qian et al. 2003). In nCPA, observations are first ordered along the stressor gradient and the point is identified where the response variable "splits" into two groups based on the greatest difference in a statistical attribute such as the mean. The changepoint is defined as the point that maximizes this difference.

A limitation of changepoint analysis is that "nCPA will identify a change point regardless of whether or not one truly exists" [emphasis added] (U.S. EPA 2010, p.54). That is, nCPA does not require an actual threshold value for the response variable. For that reason, additional analyses are required after estimating a changepoint to determine whether the selected value is, in fact, protective of the designated uses. Because changepoint analysis artificially "forces" a threshold (breakpoint) on data where an actual threshold may not even exist, it is considered here as inferior to threshold analysis using additive quantile regression (below). It is also regarded as inferior because when two groupings in the data are not apparent, the forced "artificial" breakpoint (where an actual breakpoint does not exist) is usually at a substantially

higher nutrient concentration than indicated from additive quantile regression. Several examples follow:

This analysis was among the techniques used by the state of Minnesota in setting eutrophication criteria for streams and rivers (Heiskary et al. 2013 - examples below). It generally yielded higher thresholds (changepoints) for TN and TP than additive QR. Moreover, for some stressor-response comparisions, there were broad confidence bands (indicating high uncertainty) around the changepoint (e.g., Figures A5 and A6). In various stressor-response comparisons, it was difficult to discern a changepoint "between" two groupings of the data. Moreover, an important point evident from the analysis was not considered: In the heavily disturbed Corn Belt area of southern Minnesota (South River Nutrient Ecoregion of the Minnesota Pollution Control Agency), nCPA indicated the same TP changepoint for both the percentage of tolerant macroinvertebrate taxa and the percentage of remaining "pollution-intolerant" taxa. This finding suggests that the relatively few remaining "intolerant" taxa actually were pollution-tolerant, and the pollution-sensitive ("pollution-intolerant") taxa had been eliminated (Figure A7; note the differences in the two Y axes). Based on this observation, changepoint analysis should not have been applied to the data because it was applied too late, after the streams and rivers in this major agricultural area had been seriously compromised and sensitive taxa had disappeared.

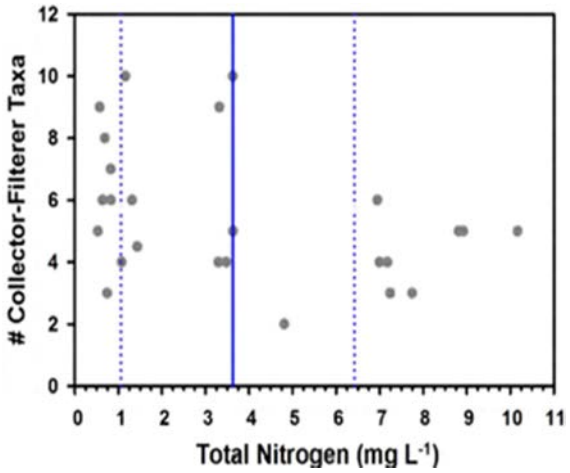


Figure A5. Relationship between TN concentration and macroinvertebrate metrics for all streams in the River Nutrient Dataset of Minnesota. Blue line = changepoint; also indicated are the 90% confidence bands. From Heiskary et al. (2013).

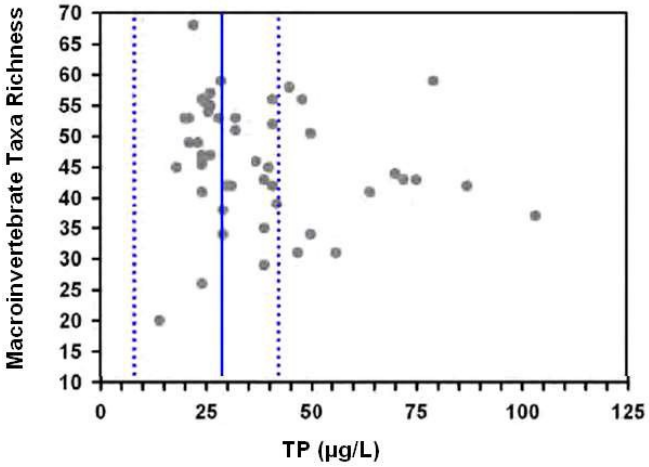


Figure A6. Relationship between TP concentration and macroinvertebrate metrics for non-wadeable streams (watershed > 500 square miles) in the North Nutrient Region of Minnesota, using a Biomonitoring dataset from the Minnesota Pollution Control Agency (MPCA). Blue line = changepoint; also indicated are the 90% confidence bands. From Heiskary et al. (2013).

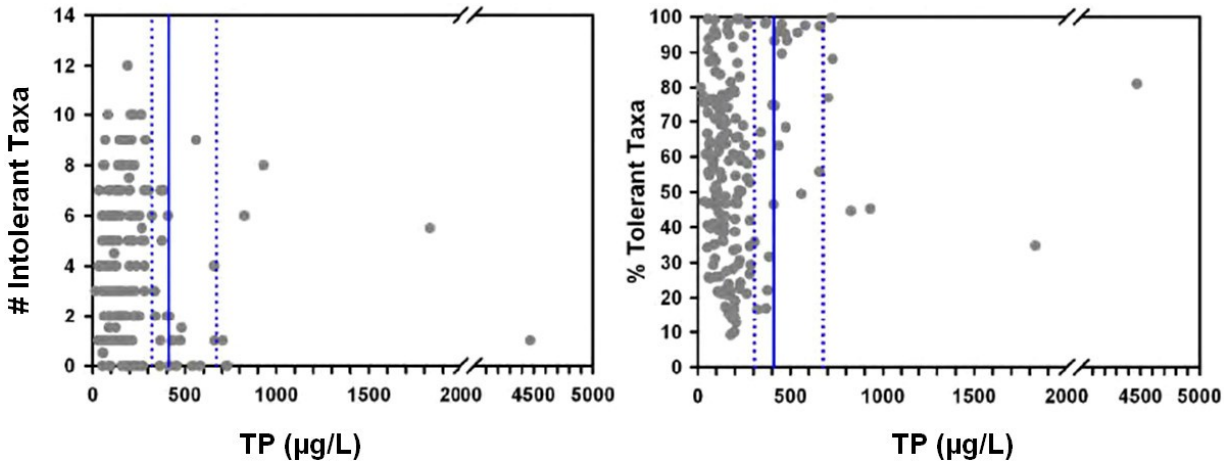


Figure A7. Relationships between TP and macroinvertebrate metrics for all streams in the South River Nutrient Region (i.e., the corn belt region) in Minnesota, using a Biomonitoring dataset from the Minnesota Pollution Control Agency (MPCA). Blue line = changepoint; also indicated are the 90% confidence bands. From Heiskary et al. (2013).

As a second example, Evans-White et al. (2009) used nCPA to estimate TN and TP thresholds for macroinvertebrate taxa numbers in 13 Central Plains streams and rivers (level III nutrient ecoregion VII; watershed size 13 hectares (32 acres or 0.05 square mile) to 8.16 million hectares (~20.2 million acres or 31,506 square miles). Benthic macroinvertebrates were collected over many years; median TN and TP concentrations collected within 30 days and 2 kilometers (1.24 miles) of a benthic macroinvertebrate sample were linked to that sample unless a point source of pollution occurred between the water quality sampling location and the macroinvertebrate sampling site.

The nCPA analysis by Evans-White et al. (2009) indicated median change point (breakpoint) values at 104-111 µg TN/L and 50 µg TP/L for total benthic macroinvertebrate taxa and benthic macroinvertebrate primary consumer taxa (Figure A8). By comparison, for wadeable streams and rivers in Wisconsin, Wang et al. (2007) reported threshold nutrient concentrations of 860 µg TN/L and 40 µg TP/L for total macroinvertebrate taxa richness, based on regression tree analysis and two-dimensional Kolmogorov-Smirnov (2DKS) threshold techniques (below). For 34 nonwadeable Wisconsin streams (41 sites), Weigel and Robertson (2007) used the technique, regression tree analysis (below), which indicated that threshold nutrient values for the same biological metric were at 1,920 µg TN/L and 150 µg TP/L.

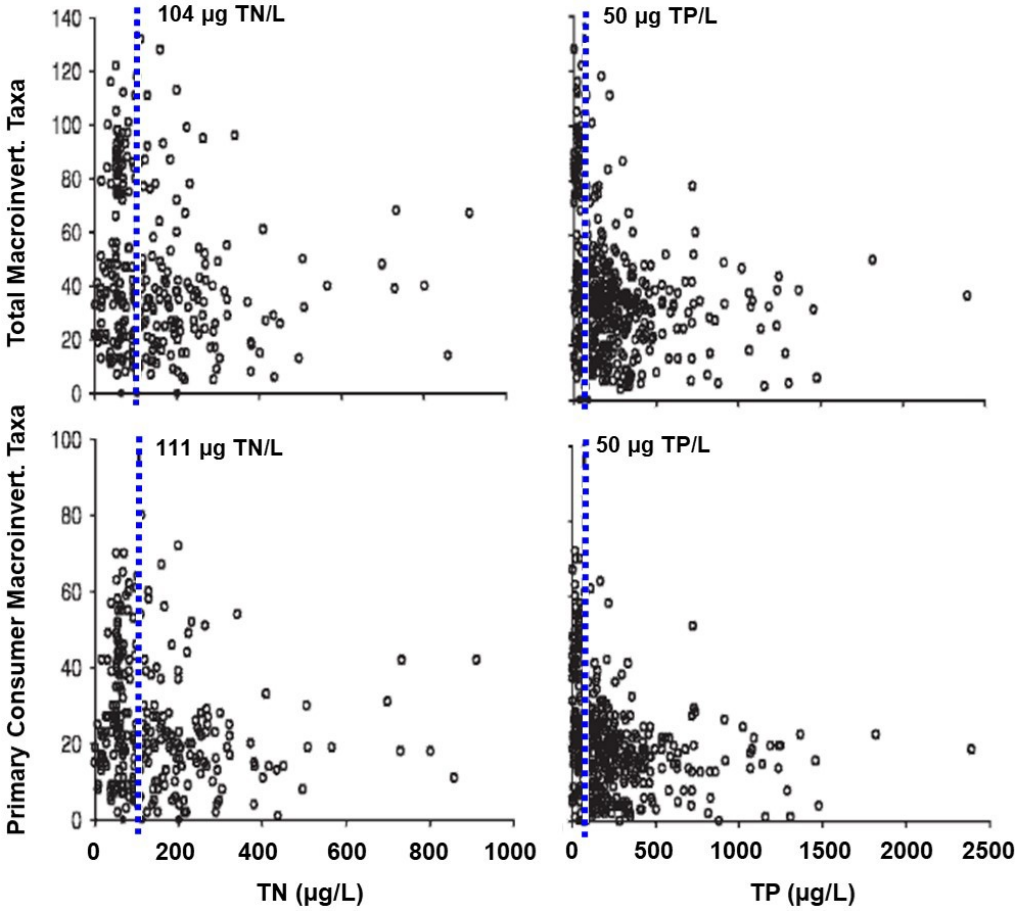


Figure A8. The total number of benthic macroinvertebrate taxa and benthic macroinvertebrate primary consumer taxa versus TN and TP concentrations, from the Central Plains streams database. Dotted vertical lines with TN and TP concentrations represent the nCPA 50% (median) threshold estimates. Modified from Evans-White et al. (2009).

As a third example of nCPA, Frey et al. (2011 - see Table A3 assessed algal, macroinvertebrate, and fish communities that best reflected impacts of nutrients along a gradient of low to high nutrient concentrations in 62 sites (1993-2006; at least six nutrient samples per year per site) within wadeable, mostly midwestern streams that reflected reference, agricultural, and urban watershed influences. These streams were within U.S. EPA nutrient ecoregions VI (Corn Belt and Northern Great Plains), VII (Mostly Glaciated Dairy Region), and VIII (Nutrient-Poor Largely Glaciated Upper Midwest and Northeast). Two "diatom ecoregions" - the Glacial North diatom ecoregion (GNE = nutrient ecoregions VII+VIII) and the Central and Western Plains diatom ecoregion (CWPE = nutrient ecoregion VI). A subgroup of 54 of the sites had biota data collected within the same year as nutrients, and were used for nutrient stressor-response analyses.

Table A3. Breakpoints between nutrients and the most significant biological metrics in the Central and Western Plains and Glacial North diatom ecoregions. From Frey et al. (2011).

Attribute (Note: All algae attributes are for diatoms)	Number of samples	Total nitrogen					Total phosphorus				
		Breakpoint concentration (mg/L)	p-value	Lower 5% confidence interval (mg/L)	Upper 5% confidence interval (mg/L)	Correlation type	Breakpoint concentration (mg/L)	p-value	Lower 5% confidence interval (mg/L)	Upper 5% confidence interval (mg/L)	Correlation type
Central and Western Plains diatom ecoregion											
Algae											
Percent pollution sensitive diatoms (Bahls)	104	NS	0.216	NS	NS	NS	0.133	**0.007**	0.064	0.140	Neg
Percent diatoms indicative of elevated concentrations of organically bound nitrogen[1]	104	4.57	**.008**	3.26	10.13	Pos	.075	**.005**	.052	.183	Pos
Percent diatoms requiring >30% DO saturation	104	3.28	**.021**	1.67	5.49	Pos	.082	**.003**	.052	.092	Neg
Percent beta-mesosaprobic diatoms[2]	104	1.67	**.041**	1.46	4.87	Pos	.133	.073	.052	.155	Neg
Abundance of diatoms preferring meso-eutrophic conditions[3]	104	4.35	.086	4.07	4.73	Neg	.070	**.016**	.053	.090	Neg
Invertebrates											
Percentage of total richness composed of bivalvia	104	8.45	**0.032**	3.45	8.80	Neg	0.189	**0.004**	0.153	0.711	Pos
Ratio of Tanytarsanii richness to midge richness	104	3.28	**.021**	1.67	10.13	Neg	.141	**.001**	.075	.192	Neg
Number of taxa in the intolerant class	104	NS	.170	NS	NS	NS	.711	**.001**	.090	.711	Neg
Abundance of Orthocladinae midges	104	4.90	**.035**	4.68	5.91	Neg	.092	**.001**	.044	.114	Neg
Fish											
Number of fish	105	4.73	**0.070**	1.24	6.95	Neg	0.070	**0.001**	0.040	0.080	Neg
Percent abundance of stonerollers	105	1.79	**.001**	1.24	4.87	Neg	.070	**.001**	.039	.070	Neg
Number of darters	105	3.51	**.007**	1.32	4.55	Neg	NS	.210	NS	NS	NS
Glaciated North diatom ecoregion											
Algae											
Percent abundance of *Achnanthidium minutissimum*	71	0.63	**0.002**	0.53	0.98	Neg	0.020	**0.001**	0.015	0.045	Neg
Richness pollution sensitive diatoms (Bahls)	71	1.32	**.001**	.56	1.54	Neg	.090	**.001**	.030	.131	Neg
Taxa richness	71	1.32	**.001**	.53	1.58	Neg	.099	**.001**	.082	.130	Neg
Percent diatom taxa generally intolerant to organically bound nitrogen[4]	71	1.20	**.008**	.43	1.48	Neg	.030	**.001**	.014	.064	Neg
Percent diatoms requiring nearly 100% DO saturation	71	1.11	**.001**	.63	1.32	Neg	.030	**.001**	.014	.060	Neg
Percent richness that are oligosaprobic[5]	71	.98	**.001**	.63	1.32	Neg	.030	**.001**	.020	.060	Neg
Percent richness diatoms preferring eutrophic conditions	71	.63	**.001**	.53	1.36	Pos	.025	**.003**	.018	.064	Pos
Invertebrates											
Richness composed of midges	72	0.66	**0.002**	0.53	1.84	Neg	0.020	**0.003**	0.017	0.082	Neg
Number of clinger taxa	72	.63	**.001**	.56	1.04	Neg	.025	**.001**	.024	.046	Neg
Richness composed of filtering-collecters	72	1.12	**.003**	.52	1.28	Neg	.025	**.001**	.018	.070	Neg
Percentage of abundance in the intolerant class	72	.56	**.002**	.56	1.20	Neg	.044	**.001**	.017	.060	Neg
Number of taxa in the intolerant class	72	.63	**.001**	.56	1.00	Neg	.024	**.001**	.020	.044	Neg
Percentage of taxa richness in the moderately tolerant class	72	.56	**.001**	.56	.98	Pos	.024	**.001**	.020	.060	Pos
Richness composed of odonates	72	.98	**.001**	.52	1.58	Neg	.046	**.001**	.017	.090	Neg
Richness composed of stoneflies	72	.63	**.001**	.53	.93	Neg	.024	**.001**	.014	.035	Neg
Richness composed of predators	72	.53	**.001**	.50	1.58	Neg	.046	**.001**	.018	.077	Neg

Table A3, cont'd.

Attribute (Note: All algae attributes are for diatoms)	Number of samples	Total nitrogen					Total phosphorus				
		Breakpoint concentration (mg/L)	p-value	Lower 5% confidence interval (mg/L)	Upper 5% confidence interval (mg/L)	Correlation type	Breakpoint concentration (mg/L)	p-value	Lower 5% confidence interval (mg/L)	Upper 5% confidence interval (mg/L)	Correlation type
Glaciated North diatom ecoregion—continued											
Fish											
Richness composed of Tanytarsanii midges	72	1.12	**0.002**	0.63	1.68	Neg	0.064	**0.001**	0.020	0.077	Neg
Percent abundance of darters	64	2.46	**.013**	1.28	2.46	Pos	.137	**.001**	.098	.170	Neg
Percent abundance of piscivorous fish	64	NS	.169	NS	NS	NS	.070	.081	.020	.130	Neg
Percent abundance of sensitive fish	64	.63	**.004**	.53	1.10	Neg	.046	**.001**	.020	.060	Neg
Percent abundance of tolerant fish	64	NS	.510	NS	NS	NS	.060	**.002**	.035	.070	Pos

[1] Obligate heterotrophs.

[2] Saprobic class: II - (02 saturation: 70–80%; BOD5: 2–4 mg/L).

[3] Beta-mesosaprobic: somewhat degraded conditions, mesotrophic-eutrophic.

[4] May be "oligotrophic" or "mesotrophic species (autotroph).

[5] Saprobic classes: I, I–II - (O2 saturation: >85%; BOD5: <2 mg/L).

In a final example of suggested numeric nutrient criteria based on nCPA, a study by Miltner (2010) of 109 small rivers and streams in Ohio indicated clear associations between nutrients, response variables benthic algal chlorophyll and DO, and macroinvertebrate and fish communities. The diel DO swing was significantly related to benthic algal chlorophyll concentrations and tree canopy cover (available light). The condition of macroinvertebrate communities was most strongly related to benthic algal chlorophyll, the minimum DO per 24-hour period, and the diel DO swing, whereas the condition of fish communities was best explained by habitat quality. For protection of still-existing high-quality waters, a changepoint was reported for benthic algae at 107 mg chlorophyll/m^2, corresponding to nutrient levels of ~440 µg dissolved inorganic N (DIN)/L and ~40 µg TP/L (or 0.44 mg DIN/L and 0.04 mg TP/L; Table A4). For management of waters that had been adversely affected by nutrient pollution, the changepoint for benthic algae was 182 mg chlorophyll/m^2, corresponding to 1,100 µg DIN/L and 100 µg TP/L (Table A4).

Table A4. Numeric nutrient criteria suggested by Miltner (2010) from a study of 109 small streams and rivers in Ohio (watershed area less than 500 square miles). "Protection" referred to streams that were evaluated as high-quality waters, whereas "management" referred to streams that had already been adversely impacted by nutrient over-enrichment.

Indicator	Protection	Management	Rationale (protection; management)
DIN (mg/l)	0.44	1.1	Change point for benthic chlorophyll; secondary mode in bootstrap samples
TP (mg/l)	0.04	0.1	Change point for benthic chlorophyll; achievable through current technology
Chl *a* (mg/m^2)	107	182	Protection of existing high-quality waters; maintain min. DO concentrations >4.0 mg/l
DO range (mg/l)	6.0	7.0	Maintain min. DO concentrations >4.0 mg/l; change point for number of EPT taxa
DO min. (mg/l)	6.0	5.0	Existing water quality standard for high-quality waters; change point for macroinvertebrate indicators
Canopy (deg open)	<40°	<40°	Change point for benthic chlorophyll

3b.3. Regression Tree Analysis (Recursive Partitioning)

Regression tree analysis has been described as ideally suited for the analysis of complex ecological data (Déath and Fabricius 2000). This statistical threshold technique can accommodate nonlinear relationships, complex factor interactions, and missing values. A regression tree explains variation in a single response variable (e.g., TP concentration) by repeatedly splitting the data into more homogeneous groups using combinations of explanatory variables. Each group is characterized by a typical value of the response variable, the number of observations in the group, and the values of the explanatory variables that define it. The regression tree is represented graphically. Regression tree analysis can be used to find breakpoints (thresholds), and to describe the relationship between a single dependent variable and various independent variables that affect the dependent variable.

As mentioned above, Weigel and Robertson (2007) used regression tree analysis to identify thresholds in biota response to TN and TP in nonwadeable Wisconsin rivers (e.g., Figure A9). While threshold nutrient values for total macroinvertebrate taxa richness were at 1,920 µg TN/L and 150 µg TP/L. Regression tree analyses also indicated that breakpoints for TP above which biota were consistently impaired ranged from 60 to 150 µg/L, depending on the biological metric. The TP breakpoint values for macroinvertebrate species and fish IBI were comparable, whereas breakpoints were lower for percent round suckers and mean pollution tolerance values (MPTVs). The TN breakpoint was ~630 µg/L for the MPTV, IBI, and percent round sucker metrics – much lower than the TN breakpoint for total macroinvertebrate taxa richness.

As a second example, Caskey et al. (2012) used regression tree analysis for a stressor-response assessment of nutrient impacts in headwater and wadeable sections of a group of 282 streams and rivers in Indiana (2005-2009; Tables A5-A7), after excluding sites known to be strongly influenced by non-nutrient variables (e.g., cold temperatures, mining operations). The objective was to assess statistically and ecologically significant relationships among stressor (TN, TP, and benthic algal chlorophyll a) versus response (macroinvertebrate and fish community) variables, including response ecologically significant relationships among stressor (TN, TP, and benthic algal chlorophyll a) versus response variables (macroinvertebrate and fish metrics), including response breakpoints. Mean nutrient concentrations ranged from 0.340 to 21.58 mg TN/L and 50-1,440 µg TP/L; mean benthic algal biomass ranged from < 1 to 719 mg chla/m^2. The combined macroinvertebrate and fish community mean TN threshold estimated from regression tree analysis was 1,030 µg TN/L and 83 µg TP/L.

These threshold TN and TP concentrations suggest mesotrophic to eutrophic stream conditions, based on Dodds and Smith (2016) (Table 2). Such conditions may have characterized the natural background in streams draining the rich soils of Indiana. On the other hand, the mean threshold for benthic algal biomass was 20.9 mg chla/m^2, borderline oligotrophic/mesotrophic conditions (Table 2), supporting the premise that the natural condition of these streams was at the low end of the mesotrophic range.

Caskey et al. (2012) noted that TN and TP concentrations in these streams spanned three orders of magnitude and that, even in the group of lowest-nutrient Indiana streams which included minimally impacted conditions, many sites were dominated by nutrient-tolerant taxa.

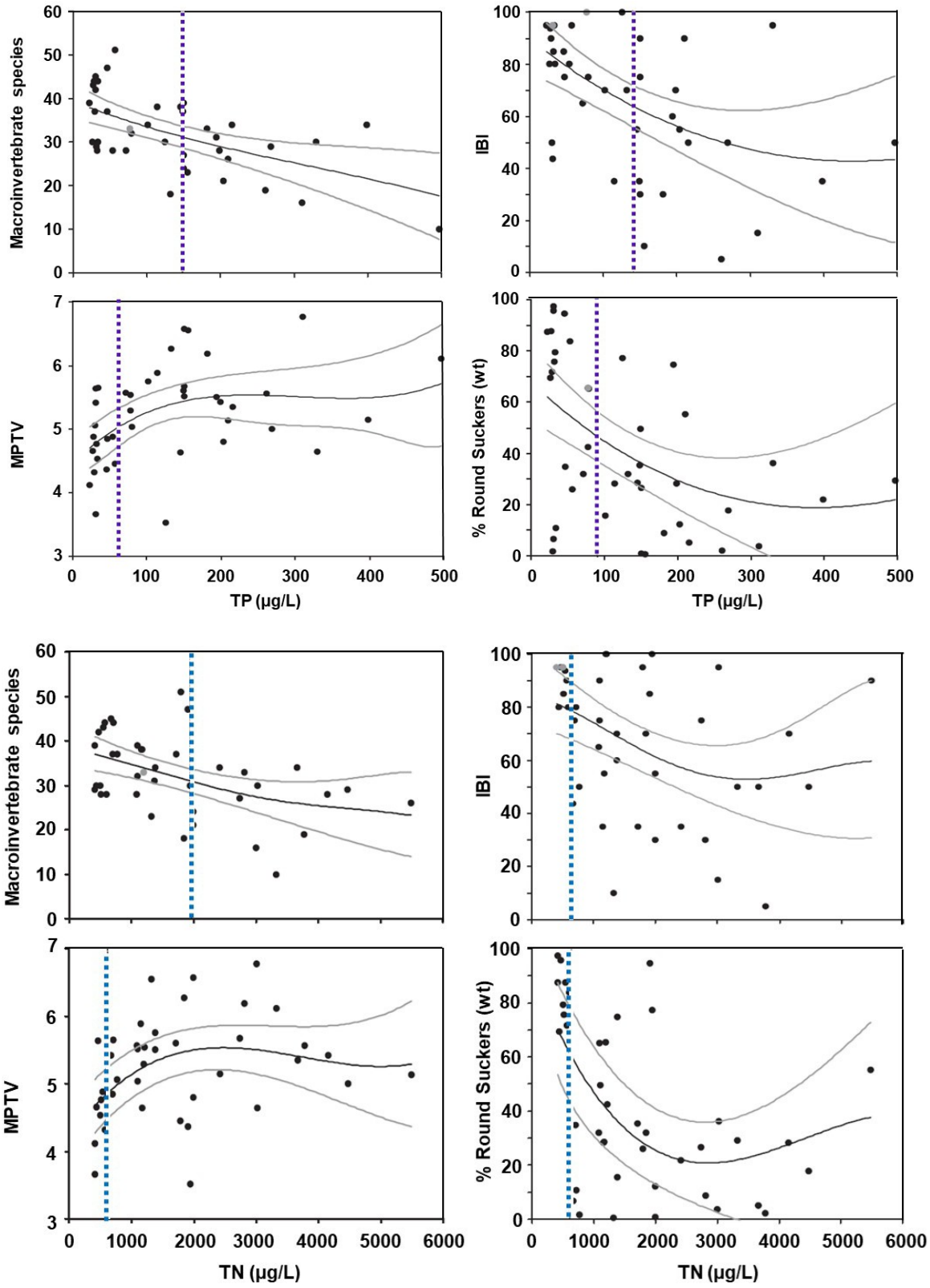


Figure A9. Upper four panels: Biological metrics vs. TP; lower four panels: Biological metrics vs. TN: Macroinvertebrate and fish measures in relation to TP or TN, including mean pollution tolerance value (MPTV) and fish index of biotic integrity (IBI). LOESS smoothing lines (black) suggest trends with 95% confidence limits (gray lines). Vertical line depicts breakpoints from regression tree analysis. Gray dots indicate two sites with the same values. Modified from Weigel and Robertson (2007).

Table A5. Indiana stream study - example of findings for TN, TP, and benthic microalgal (periphyton) biomass as chlorophyll a data ordered into quantiles - summary statistics for the stressor variables for headwater and larger, wadeable sites (5-year study). Modified from Caskey et al. (2012).

Statistic	Total nitrogen (mg/L)	Total phosphorus (mg/L)	Periphyton CHL*a* (mg/m²)
Headwater sites (≤52 km²) (n=126)			
Minimum	0.343	0.050	0.947
10th percentile	.564	.050	9.18
25th percentile	.944	.055	28.1
Median	2.12	.093	69.7
75th percentile	3.66	.140	143
90th percentile	5.92	.292	230
Maximum	21.6	1.44	629
Wadeable sites (>52 km² ≤2,590 km²) (n=156)			
Minimum	0.340	0.050	0.383
10th percentile	.768	.057	7.29
25th percentile	1.22	.078	19.2
Median	2.31	.110	44.7
75th percentile	3.52	.157	115
90th percentile	4.85	.234	228
Maximum	10.0	1.24	719

Indicates sites with "Low" concentrations less than or equal to the 10th percentile.

Indicates sites with "High" concentrations greater than the 75th percentile.

Table A6. Breakpoint concentrations for TN, TP, and benthic chlorophyll a (Peri = periphyton or benthic algae) found in this study of Indiana headwater and larger wadeable streams, including comparison with the trophic categories described by Dodds et al. (1998). From Caskey et al. (2012).

Causal variable	Lowest breakpoint	Mean low breakpoint	Dodd's oligotrophic-mesotrophic	Dodd's mesotrophic-eutrophic
TN (mg/L)	0.314	1.03	0.700	1.50
TP (mg/L)	.050	.083	.025	.075
Peri CHL*a* (mg/m²)	3.49	20.9	20.0	70.0

Table A7. Summary of the most similar breakpoint-threshold concentrations from regression tree analysis, versus quantile-threshold (QR) concentrations from the additive QR analysis for Indiana headwater and wadeable streams (Peri = periphyton or benthic algae). From Caskey et al. (2012).

Causal variable	Community	Response variable	Number of occurrences	Breakpoint concentration	p-value	Quantile threshold concentration
			Headwater			
TN (mg/L)	Invertebrate	*Ablabesmyia mallochi*	25	0.656	0.036	0.863
TP (mg/L)	Invertebrate	Percent intolerant species	117	.057	.001	.073
			Wadable			
TP (mg/L)	Invertebrate	Margalef Index	156	0.153	0.031	0.121
TP (mg/L)	Fish	Tolerant percent	156	.144	.001	.157
Peri CHL*a* (mg/m²)	Invertebrate	EPT	148	27.2	.001	14.0

Protective threshold TN and TP concentrations, such as those reported by Caskey et al. (2012), are not commonly reported for highly nutrient-enriched streams of the Midwest (e.g., Illinois, Indiana, and Ohio) because minimally impacted conditions can seldom be found and, therefore, are usually not included. Thus, nutrient thresholds for adverse impacts on biota more typically are reported to be three- to five-fold higher than thresholds for streams in less nutrient-enriched waters (Frey et al. 2011). The reason for this difference is that the heavily nutrient-impacted streams have only pollution-tolerant taxa remaining. Similarly as for the heavily impacted corn belt streams of southern Minnesota, it is too late to attempt to find a threshold for sensitive taxa unless true minimally impacted waters can be found; very few sensitive taxa, low in number, have survived in most waters. The high "nutrient thresholds" indicated from most analyses are nutrient levels which adversely impact the dominant, tolerant species that remain.

Significant zero crossing (SiZer) is defined as a model that explores thresholds in relationships between a single stressor (e.g, TN or TP) and a single (univariate) response variable (Wagenhoff et al. 2012). It can be used to characterize multiple ecological thresholds based on significant changes in the slope of a relationship between response (dependent) variables and causal (predictor or independent) variables (Chaudhuri and Marron 1999, Clements et al. 2010, Daily et al. 2012). This technique uses smoothing and identifies thresholds based on the derivatives of the smoothed curve. The threshold is the point at which a derivative changes from being statistically significant (different from zero) to not being significant (not different from zero) or vice versa; thus, multiple thresholds across the range of the stressor can be identified (Sonderegger et al. 2009, Wagenhoff et al. 2012).

The changes in the derivative as a function of the predictor variable (e.g., TN or TP) are displayed in a single "user-friendly" map (Clements et al. 2010). The map, actually a graph, indicates potential thresholds across the stressor gradient at different levels of resolution (note that it is up to the analyst to decide the appropriate level of resolution, or bandwidth, from an ecological perspective).

SiZer is described as "especially appropriate in situations where incorrectly identifying a threshold has serious consequences for protecting aquatic resources (e.g., overestimating a critical threshold concentration for a contaminant)" (Clements et al. 2010, p.1022). Since the approach is only exploratory, however, it does not involve testing for the statistical significance of the thresholds, and does not provide measures of uncertainty (confidence intervals) around the threshold estimates.

SiZer and other statistical threshold analyses were applied to a set of streams draining watersheds with 0 to 90% agricultural land use, to assess whether land use nonlinearly affected nutrient concentrations and, if so, whether breakpoints could be detected (Xenopoulos et al. 2013). Nonlinearity was strongly detected; breakpoints ranged from 46 to 70 µg TP/L for several key ecosystem processes. SiZer has also been applied to quantify potential recovery thresholds of metal-sensitive macroinvertebrate taxa after restoration of a river affected by mining (Clements et al. 2010), and could be analogously applied to rivers affected by nutrient pollution (but see below).

Two-dimensional Kolmogorov-Smirnov test (2DKS) is a nonparametric method for identifying breakpoints in variance for two variables (bivariate data). It allows analysis of different types of data wherein a linear relationship is not expected, and any functional relationship between the variables breaks down over some range of values of one of the variables, so that high variance in the other variable is expected across that range (Garvey et al. 1998). The 2DKS test can detect non-random patterns in bivariate data. If the independent variable (TN or TP) controls biotic responses only beyond some critical level, the 2DKS test can detect and estimate the threshold level. More commonly, the variance in a dependent variable such as benthic algal biomass may not be constrained at some fixed threshold value of TN or TP. Instead, the bivariate distributions can vary in complex ways. Because the 2DKS test has no prior assumptions about the data distribution, it is robust to non-random patterns in bivariate distributions (Garvey et al. 1998 and references therein).

A 2DKS was used to assess how water-column nutrients were linked to benthic algal biomass in temperate streams (nearly 300 sampling periods), and in a separate dataset for 260 temperate stream sites across the U.S. (Dodds et al. 2002). A 2DKS test indicated significant breakpoints of 27 µg TP/L and 515 µg TN/L, above which mean benthic algal chlorophyll was substantially higher (Dodds et al. 2002, 2006) (also see Appendix - section C below).

C. Derivation of Numeric Nutrient Criteria Using Thresholds - Which Method?, and How Do the Methods Compare?

Selection of the threshold analysis method for use should be based on the amount of data (U.S. EPA 2000a,b), the observed data distributions (Brenden et al. 2008), and knowledge of the ecological dynamics and potential threshold patterns of the waterbodies being considered (Qian and Cuffney 2012). It is worth emphasizing again that statistical threshold analysis requires considerable expertise in order to avoid pitfalls that cause the analysis to yield incorrect information. As Wagenhoff et al. (2012, pp. 10 and 30) noted,

> Lack of statistical understanding or simply a failure to report model parameters is a widespread issue in the [published science] literature

> concerned with…threshold detection. Many authors do not justify why a particular threshold model was chosen and also rarely seem to pay attention to…the statistical validity of their modeling approaches. In some cases authors neither test for statistical significance nor give any measure of uncertainty around the estimate, which is central to [these] techniques…. The pattern of a threshold response is likely specific to each driver-response pair, making it necessary to choose and apply the most appropriate model and method for each case. Appropriate use of [these] methods also requires an understanding of the statistical assumptions underlying the tests….

Relatively few comparisons of these techniques applied to the same dataset and response variables are available, as well (Dodds et al. 2010, Wagenhoff et al. 2012, Evans-White et al. 2014). Brenden et al. (2008) compared five different approaches to threshold detection for three simulated datasets, each of which showed a threshold but differed in the threshold pattern. No one method was identified as superior to the others.

Dodds et al. (2010) examined thresholds in macroinvertebrate richness across stream TP concentrations using breakpoint regression, CFD, quantile regression tree, nCPA, 2DKS, and SiZer. Threshold estimates in that study varied up to three-fold depending on the statistical method used. Breakpoint regression, 2DKS, and SiZer yielded the highest, least protective threshold TP concentrations.

Xenopoulos et al. (2013) compared breakpoint regression, 2DKS, and SiZer to test for nonlinearity and ecological thresholds for some temperate streams draining watersheds that ranged from 0 to more than 90% agricultural land use, to assess whether land use altered in-stream nutrients and ecosystem function (decomposition, benthic algal biodiversity). As mentioned (p.88), breakpoints ranged from 46 to 70 µg TP/L for some key ecosystem processes.

Daily et al. (2012) evaluated the performance of commonly used statistical threshold analyses (piecewise quantile regression, nCPA, and SiZer) applied to stream benthic macroinvertebrate communities. False detection of ecological thresholds were influenced by sample size, the rate of change, and the frequency of observations across the environmental gradient. The relative importance of these factors varied among statistical methods. For example, false threshold detection rates were influenced primarily by user-selected parameters for piecewise quantile regression and SiZer, and secondarily by sample size. If the sample size was small (< 50), SiZer appeared to minimize false threshold detection more than piecewise quantile regression analysis. The authors concluded that the choice of statistical methods for threshold detection should be matched to experimental and environmental constraints insofar as possible, in order to avoid false detection of ecological thresholds and spurious inferences.

Considering the above information, a major recommendation by Wagenhoff et al. (2012) - use of multiple statistical methods insofar as possible - is strongly endorsed here, in combination with a concluding recommendation by Dodds et al. (2010): The numeric criteria indicated by the most protective analyses should be adopted as a precautionary approach, given the scientific uncertainty about the efficacy and the lag period in recovery of aquatic ecosystems once a threshold point has been crossed.