New Mexico Nutrient Thresholds for Perennial Wadeable Streams

Final Report

August 21, 2015

Prepared by Tetra Tech, Inc.
in cooperation with the
New Mexico Environment Department
and the
U.S. EPA Region 6 and the N-STEPS Program

New Mexico Nutrient Thresholds for Perennial Wadeable Streams – Final Report

August 21, 2015

Prepared by Tetra Tech, Inc.

in cooperation with the

New Mexico Environment Department

and the

U.S. EPA Region 6 and the N-STEPS Program

Author: Ben Jessup, Tetra Tech, Inc.

Workgroup: Seva Joseph (NMED), Bryan Dail (NMED), Lynette Guevara (NMED), Shelly Lemon (NMED), Scott Murray (NMED), Forrest John (EPA R6), Jacques Oliver (EPA OW), Lester Yuan (EPA OW), Christopher Patrick (EPA OW), Michelle Maier (EPA OW), Michael Paul (Tetra Tech)

This work was supported by EPA's Office of Water, Office of Science and Technology, through the Nutrient Scientific Exchange and Partnership System (N-STEPS) administered by EPA's National Nutrient Criteria Program.

Executive Summary

The State of New Mexico currently has a narrative criterion for plant nutrients. This criterion is used to identify nutrient conditions that contribute to production of undesirable or nuisance aquatic life. Narrative criteria must be translated to numeric nutrient thresholds for consistent implementation in impairment determinations, NPDES permit limits, and TMDL budgets. Therefore, the Surface Water Quality Bureau (SWQB) of the New Mexico Environment Department (NMED) is refining nutrient threshold values with regional data, reference conditions, links between cause and response variables, and verified classification systems.

The United States Environmental Protection Agency (USEPA) nutrient criteria guidance recommends that criteria be derived for primary causal variables total nitrogen (TN) and total phosphorus (TP). The analysis presented in this report uses concurrently measured causal and response variables, including diatom and benthic macroinvertebrate assemblages, dissolved oxygen (DO) and chlorophyll *a* (chl-a) concentrations. The approach uses reference conditions and stressor-response relationships to derive numeric nutrient thresholds for application of the narrative criterion. This document describes the nutrient threshold development process and existing nutrient conditions in New Mexico streams across the landscape, in relation to the stressor gradient and aquatic life uses.

The pathways by which nutrients affect aquatic life conditions are the basis for the stressor-response analyses. The focus was on TN and TP effects on diatoms and benthic macroinvertebrates through direct and indirect pathways including chl-a and DO. Chl-a is a proxy for algal biomass, which in turn is a proxy for algal productivity. Algal productivity is the link between the nutrient stressors and the impacts to designated uses (e.g., DO consumption, nuisance to humans, displacement of other desirable algae, etc.). The linkages and relationships defined in the conceptual model and analyzed in this study include:

1. Increases in nutrients result in increases in chl-a
2. Increases in chl-a result in changes in DO dynamics
3. Increases in nutrients result in changes in diatom metrics
4. Changes in DO dynamics result in changes in macroinvertebrate metrics

The analytical methods consisted of two major categories: reference condition distributions and stressor-response relationships. The reference condition approach included identification of minimally disturbed sites, classification of the sites, and description of reference condition based on characteristics in minimally disturbed sites in each site class. Candidate thresholds were derived from distributions of nutrient concentrations in reference streams. In this study, reference streams represent least disturbed conditions with ecological integrity expected for a region. As such, the nutrient conditions in reference streams are the best estimate concentrations due to “*natural causes*” as they are referred to in the narrative criteria. Thresholds for a particular stressor of interest (e.g., TN or TP) were derived by selecting a

representative value from the distribution found among reference sites. Values greater than those observed in the reference conditions are likely to be outside of the reference range, indicating the presence of *"nutrients from other than natural causes"* and possible threats to aquatic life designated uses. The threshold value defines expectations for all streams in a site class.

Stressor-response approaches refer to analytical techniques that derive candidate thresholds by defining the relationships between response variables and nutrient concentrations. Response variables included chl-a and DO, as well as biological metrics that are known to respond to nutrient stressors. Benthic macroinvertebrate and diatom metrics represent the relative integrity of the community at a site. Maintaining metric conditions that are similar to conditions observed in reference sites is a reliable indicator that designated aquatic life uses are being protected.

The statistical techniques for relating stressors and responses included correlation analysis, regression interpolation, and change-point analysis. Each of these techniques has strengths and limitations that inform the use of resulting candidate thresholds. The reference condition approach was emphasized and evidence from the stressor-response relationships further supported threshold selection.

Data were collected through NMED and national monitoring programs within New Mexico and in the immediate surrounding areas, including the National Rivers and Streams Assessment (NRSA) and the Wadeable Streams Assessment and Environmental Monitoring and Assessment program (WSA & EMAP, respectively). A GIS analysis of sites and their catchments was conducted to characterize environmental conditions. All data were compiled in a relational database following a Quality Assurance Project Plan (QAPP) (Tetra Tech 2011a, 2012). Screening sites for data integrity and completeness resulted in 663 sites with nutrient data in one or more samples collected between 1990 and 2012. Other types of data (diel DO, chl-a, macroinvertebrates, and diatoms) were available for subsets of those sites. All data were screened for outlier values and nutrient values were standardized to common detection limits.

The reference site analysis and disturbance gradient designations of 542 sites resulted in 20% of sites identified as least disturbed reference sites. Another 11% were designated as near-reference. The reference and near-reference sites were used to determine site classes based on nutrient conditions. For nitrogen, concentrations were associated with land slope, and three nutrient classes were identified as TN Flat, TN Moderate, and TN Steep sites. For phosphorus, soil TP and volcanic geology were important in addition to land slope, resulting in three different nutrient classes: TP High-Volcanic, TP Flat-Moderate, and TP Steep. In general, nutrient concentrations are higher in flatter landscapes. Frequency distributions of nutrient conditions in reference and near-reference sites were used to derive candidate thresholds. Median nutrient values within sites and the 90th quantile of median values across sites within site classes were the basis for the candidate thresholds.

Correlation and other multivariate techniques supported the major linkages between nutrients, chl-a, DO, diatoms, and macroinvertebrates. Chl-a relationships supported causal linkages between chl-a, nutrients, and DO but were too weak and variable to derive candidate thresholds. The indirect correlations between nutrients and macroinvertebrates were as strong as more direct correlations between nutrients and DO, or DO and macroinvertebrates. Although chl-a and DO were assumed to be intermediate in the pathway of nutrient-macroinvertebrate effects, indirect nutrient-macroinvertebrate relationships were also defined. The direct nutrient-diatom relationship was evident, therefore intermediate links were not analyzed. Co-varying conditions (e.g., conductivity with nitrogen) that might contribute to the relationships between nutrients and responses were also found. These were noted for further consideration, but could not be effectively factored out of the stressor-response analyses.

Regression interpolations and change-point analysis for macroinvertebrate and diatom metrics and DO in response to nutrient concentrations resulted in multiple candidate thresholds in each site class. Candidate thresholds that did not pass criteria of significance or corroboration of analytical indicators were eliminated. Candidate thresholds from stressor-response analyses defined a range of values around the reference frequency distribution thresholds.

For each nutrient and site class, candidate thresholds from all analyses were quantified, weighting the reference distribution results as primary because they were direct evidence of nutrient conditions in minimally disturbed systems. The stressor-response results typically bracketed the reference distribution results. The range of candidate thresholds are presented in cumulative distribution function (CDF) curves, including confidence intervals around the primary thresholds as well as ranges of alternative thresholds derived from stressor-response analyses. In general, the primary thresholds were comparable to the 50^{th} - 70^{th} quantile of the candidate thresholds derived from stressor-response analysis based on the CDF curves.

		TN Flat	TN Moderate	TN Steep
TN	Reference 90^{th} quantile	0.69 mg/L	0.42 mg/L	0.30 mg/L
	90% confidence interval	0.62 – 0.85	0.38 – 0.51	0.26 – 0.34
	Stressor-response median	0.52 mg/L	0.33 mg/L	0.26 mg/L
		TP High-Volcanic	TP Flat-Moderate	TP Steep
TP	Reference 90^{th} quantile	0.105 mg/L	0.061 mg/L	0.030 mg/L
	90% confidence interval	0.089 – 0.114	0.051 – 0.069	0.016 – 0.053
	Stressor-response median	0.067 mg/L	0.066 mg/L	0.029 mg/L

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1 Purpose

While a few streams have segment specific numeric criteria for total phosphorus, the State of New Mexico currently has no general numeric criteria for nutrients. The narrative criterion in the State of New Mexico Standards for Interstate and Intrastate Surface Waters found at § 20.6.4.13 NMAC provides that (NMWQCC 2011):

> *Plant nutrients from other than natural causes shall not be present in concentrations which will produce undesirable aquatic life or result in a dominance of nuisance species in surface waters of the state.*

The narrative nutrient criterion can be difficult to implement because the relationships between nutrient levels and impairment of designated uses are not quantitatively defined, and distinguishing nutrients from "other than natural causes" is difficult (NMED/SWQB 2008). Therefore, the Surface Water Quality Bureau (SWQB) has developed nutrient assessment protocols for streams and is planning to refine the protocols and nutrient threshold values with regional data, links between cause and response variables and verified classification systems.

Towards the implementation of this narrative criterion, New Mexico has adopted an assessment method applicable to wadeable perennial streams that evaluates nutrient impairment for the purpose of Clean Water Act § 303(d) listing and TMDL development. The wadeable stream assessment utilizes a tiered weight-of-evidence approach that includes dissolved oxygen, pH, total nitrogen (TN), total phosphorus (TP), qualitative algae coverage, periphyton coverage, and anaerobic conditions observations at the screening level, and quantitative measures of both stressor and response variables using either a threshold or, in unique cases, reference-based approach. The State's use of the nutrient assessment protocol has resulted in the listing of 54 assessment units (i.e., stream reaches) and the development of 31 EPA-approved nutrient TMDLs for TN and/or TP. Although successfully implemented, the NMED assessment protocols for nutrients, is based on thresholds derived from frequency distribution curves and were never linked to undesirable responses or use impairment.

The United States Environmental Protection Agency (USEPA) nutrient criteria guidance recommends that criteria be derived for primary causal variables TN and TP and primary response variables chlorophyll-a (chl-a) and clarity. EPA's guidance supports modeling nutrient stressors with response variables as a way of deriving TN, TP, or chl-a thresholds and suggests using responses such as dissolved oxygen, trophic state indices, and biocriteria (USEPA 2000). EPA recommends three methods to establish nutrient criteria (USEPA 2000): a reference-based approach, a stressor-response approach, and literature-derived values.

The NMED proposed a stepwise threshold development approach as described in Empirical Approaches for Nutrient Criteria Derivation (USEPA 2009). This approach includes five steps, (1) Selecting and Evaluating Data; (2) Assessing the Strength of the Cause-Effect Relationship; (3)

Analyzing Data; (4) Evaluating Estimated Stressor-Response relationship; and (5) Evaluating Candidate Stressor-Response Criteria. Toward this end, New Mexico SWQB, working with Tetra Tech, completed a preliminary analysis (a Proof of Concept) that undertook Step 1 and started the analysis for Steps 2 and 3. This report describes the ongoing analyses up to Step 5. The goal at this time is to propose revised numeric thresholds for New Mexico's narrative nutrient water quality standard. The analysis described herein uses concurrently measured causal and response variables including nutrients and other related water quality parameters, as well as biological data, i.e., algal and benthic macroinvertebrate community composition and chl-a concentration. The development approach for nutrient thresholds will use reference conditions and stressor-response relationships to derive numeric thresholds (USEPA 2009).

This report describes existing nutrient conditions in New Mexico streams across the landscape, in relation to the stressor gradient and aquatic life uses. With these descriptions in mind, the purpose of this document is to describe the intent, methods, and results of nutrient threshold development for wadeable streams in New Mexico

1.2 Background

Nutrient threshold development for TN and TP values in perennial, wadeable streams in New Mexico has taken place in three steps, thus far. First, the EPA compiled nutrient data from the national nutrient dataset, divided it by waterbody type, grouped it into nutrient ecoregions, and calculated the 25th percentiles for each aggregate and Level III ecoregion (Table 1). EPA published the Clean Water Act 304(a) recommended water quality criteria for TN and TP to help states and tribes reduce problems associated with excess nutrients in waterbodies in specific areas of the country (USEPA 2000).

Table 1. Draft Level III Ecoregion Nutrient Criteria for streams (mg/L), calculated using 25th percentile by EPA procedures, draft Ecoregion Nutrient Criteria (USEPA 2000).

	21-Southern Rockies	23-AZ/NM Mountains	22-AZ/NM Plateau	24-Chihua-huan Desert	26-SW Tablelands
TN	0.04	0.12	0.085	0.543	0.26
TP	0.006	0.011	0.015	0.018	0.025

Refinement of the ecoregional nutrient thresholds for New Mexico was conducted by Evan Hornig, a USGS employee assisting states in EPA Region 6 with development of nutrient thresholds. Hornig used regional nutrient data from EPA's Storage and Retrieval System (STORET), the U.S. Geological Survey (USGS), and the Surface Water Quality Bureau (SWQB) to create a regional dataset for New Mexico. The revised threshold values were calculated based on EPA procedures and the median for each Level III ecoregion (Table 2).

Table 2. Ecoregional nutrient thresholds for streams (mg/L), calculated using regional data, the 50th percentile and EPA procedures (Evan Hornig, unpublished data 2003)

	21-Southern Rockies	23-AZ/NM Mountains	22-AZ/NM Plateau	24-Chihuahuan Desert	26-SW Tablelands
TN	0.30	0.32	0.42	0.64	0.54
TP	0.025	0.020	0.070	0.062	0.025

The third round of analysis was conducted by SWQB to produce nutrient threshold values for streams based on ecoregion and designated aquatic life use. For this analysis, TP, total Kjeldahl nitrogen (TKN), and nitrate plus nitrite (NO_3NO_2) data from the National Nutrient Dataset (1990-1997) was combined with Archival STORET data from 1998 and 1999-2006 data from the SWQB in-house database. SWQB recognized site classes based on Level 3 ecoregions and cold, warm, and transitional temperature regimes as applied for designated Aquatic Life Uses. The threshold values (Table 3) were derived from median values for all the data in each classification. The process that SWQB used for this analysis is detailed in Appendix A.

Table 3. Ecoregional nutrient and aquatic life use thresholds for streams (mg/L), using regional data and the 50th percentile (NMED/SWQB 2013).

	Southern Rockies (21)		AZ/NM Plateau** (22)		AZ/NM Mountains (23)		Chihuahuan Desert** (24)	SW Tablelands (26)		
TN	0.25		0.35		0.25		0.53	0.38		
TP	0.02		0.05		0.02		0.04	0.03		
ALU*	CW	T/WW (volcanic***)	CW	T/WW	CW	T/WW	T/WW	CW	T	WW
TN	0.25	0.25	0.28	0.48	0.25	0.29	0.53	0.25	0.38	0.45
TP	0.02	0.02 (0.05)	0.04	0.09	0.02	0.05	0.04	0.02	0.03	0.03

<u>NOTES</u>: * ALU = designated aquatic life use of the assessment unit
CW – streams with only coldwater uses (high quality coldwater or coldwater)
T – transitional streams with marginal coldwater, coolwater, or both cold and warmwater uses
WW – streams with only warmwater uses (warmwater or marginal warmwater)

** Because of the limited area and number of sites in the Madrean Archipelago (79) and Colorado Plateau (20) ecoregions, these data where grouped with the most similar ecoregions; the Madrean Archipelago with the Chihahuan Desert and the Colorado Plateau with the Arizona/New Mexico Plateau. The Western High Plains (25) had no stream data as the only surface waters are playas, therefore this protocol does not apply to this ecoregion.

*** The volcanic threshold is applicable to Level IV ecoregions 21g and 21h as well as 21j in the Jemez Mountains

The relationship between nuisance algal growth and nutrient enrichment in stream systems has been well documented in the literature (Van Nieuwenhuyse and Jones 1996; Dodds et al. 1997; Chetelat et al. 1999, Suplee et al. 2009, Stevenson et al. 2006). The NMED assessment protocols (2013) currently include procedures for assessing algal growth using visual assessments and measures of chl-a and at least 72 hours of continuous DO and pH monitoring. Threshold values are applied via New Mexico's listing methodology in a tiered, weight-of-evidence approach.

Exceedance of the TN or TP thresholds and confirmation of potential impairment from at least one response indicator (algal abundance, dissolved oxygen, or presence of anoxic sediments) during the screening level assessment triggers additional data collection and a complete nutrient assessment. For the screening level, NMED assumes that high rates of primary production can cause DO super-saturation and high pH during the day. Impairment is suspected if DO saturation readings are above 120% and/or pH values are above the appropriate aquatic life criterion (i.e., pH > 8.8 for high quality cold and coldwater uses or pH > 9.0 for marginal cold, cool, warm, and marginal warmwater uses). In the complete nutrient assessment, DO and pH data are collected using multi-parameter, continuous recording devices to observe diel fluctuations, as opposed to the "snapshot" that one-time, grab data provide. Because algal biomass above nuisance levels often produces large diel fluctuations in DO and pH (Mabe 2007), DO concentration, percent local DO saturation, and pH are used as indicators of nuisance levels of algal biomass. For algal biomass, complete nutrient assessments include collection of a benthic periphyton sample, analysis of chl-a concentration and comparison of the concentration to thresholds (Table 4). When multiple nutrient concentration measures are taken within an assessment unit, more than 10% of TN and/or TP measurements must exceed the threshold, and one or more response variables must be present, before impairment is determined during the complete nutrient assessment (NMED/SWQB 2013). The assessment units are defined by NMED and represent waters with assumed homogenous water quality, such as a stream segment between major tributaries.

Table 4. Chl-a Level III Ecoregional Threshold Values in $\mu g/cm^2$. Reproduced from NMED (2011).

21-Southern Rockies	20/22-AZ/NM Plateau	23-AZ/NM Mountains	24/79-Chihuahuan Desert	25/26-SW Tablelands
3.9 – 5.5	7.4 – 7.8	5.8 – 11.0	16.5 – 17.5	8.2 – 14.0

NMED observed that the current nutrient thresholds were frequently exceeded at sites with little human activities in the watershed and therefore did not provide an effective filter for identifying impairment from "…*other than natural causes*…" (20.6.4.13(E) NMAC). Since 2002, exceedances of more than 15% of samples at each site were noted in high percentages of sites (43-100%, commonly >80%) in each ecoregion (Seva Joseph, personal communication). NMED has new data for analysis, including data from a broader region, and a new analytical approach based on reference conditions and stressor-response analysis. Therefore, NMED initiated the process for revising the thresholds starting with a data gathering and exploration phase, the Proof of Concept for nutrient threshold development (Tetra Tech 2011b). Variables were selected for analysis, data were assembled, and characteristics of these data were explored. Data regarding nutrients, water chemistry, physical habitat conditions, site characteristics, and response variables were compiled from multiple sources. Data exploration consisted of preliminary analysis using techniques intended for final analysis once the datasets are fully assembled. The preliminary analyses and data visualization tools were used to select variables to appropriately quantify the stressors (i.e., excess nutrients) and the responses. Collaborators on the several phases of nutrient threshold development have included NMED, USEPA Region 6, the Nutrient Scientific Technical Exchange Partnership and Support (N-STEPS) program run through the U.S. EPA's Office of Water Nutrient Criteria program, run through the Health and Ecological Criteria Division (HECD), and the consulting firm Tetra Tech, Inc.

1.3 Linking NM's Narrative Criterion to Nutrient Stressors

Nutrients occur in streams naturally and can be greatly increased due to human activity. In this study, the focus was on nitrogen and phosphorus because these nutrients (or other non-nutrient factors) typically limit or enhance primary production and are readily measured. Other nutrients are usually only required in trace amounts for plant growth and rarely limit production. Therefore, increases in nutrients other than nitrogen and phosphorus might be evident with increased human disturbance, but they are not suspected of causing changes in the primary producers.

Human activities can cause increases in nutrient concentrations in streams through a variety of pathways. These include, but are not limited to, fertilizer application, soil and vegetative disturbance, partial treatment of municipal or residential wastewater, and animal production. These sources are presumed to be related to general classes of land cover, including agriculture and residential/urban development and to specific human activities such as wastewater treatment.

Increases in major nutrients are often associated with increases in other pollutants and stressors. Nutrients may be associated with turbidity and Total Suspended Solids (TSS). Suspended sediments, in turn, have been associated with metrics of the benthic macroinvertebrate assemblage (Jessup et al. 2010). The interaction of multiple stressors can cause amplified or buffered effects on responding organisms. This phenomenon was explored in this analysis, though the emphasis remains on the interaction between major nutrients, secondary stressors, and biotic responses.

Forms of nitrogen and phosphorus vary depending on the conditions in their environment. Some forms, such as ammonium ions (NH_4^+), nitrate ions (NO_3^-), and orthophosphate (PO_4^{-3}), are more accessible for uptake by plants. Typical nutrient measures from stream samples include nitrate (NO_3), nitrite (NO_2), combined NO_3NO_2, ammonium (NH_4), total Kjeldahl N (TKN), total nitrogen (TN), orthophosphate, and total phosphorus (TP). The most common measures are NO_3NO_2, TKN, TN, and TP. TN can be calculated from TKN + NO_3 + NO_2. The best indication of potential nutrient availability is the sum of all forms, or TN and TP.

Protection of aquatic life uses is required by the Clean Water Act and the key reason for establishing nutrient thresholds. The pathways by which nutrient concentrations affect aquatic life conditions are complex, as suggested in conceptual models (e.g., EPA 2010, EPA 2012) and literature supporting linkages along the pathways. The basic relationships are described here and form the basis of our rationale for using the selected response measures in the analyses.

Nutrient impaired waters can cause problems that range from annoyances to serious health concerns (Dodds and Welch 2000). In streams, gross primary production is effected by nutrient concentrations, especially phosphorus (Mulholland 2001). The primary producers include periphyton and aquatic macrophytes. Periphyton (including diatoms) is ubiquitous in streams and can be sampled consistently. They are therefore potential indicators of nutrient conditions. Periphyton species are responsive to stressors other than nutrients, especially in the West (Stevenson et al. 2008), but these confounding factors (e.g., canopy cover, flow, turbidity) may be recognized and perhaps even factored out of descriptive stressor-response relationships.

Periphyton produce oxygen when photosynthesizing, but can deplete oxygen as well, during periods of respiration and when microbes respire in the decay of excessive periphyton,. Production and respiration in streams are can be assessed through examination of the diel DO range (Mulholland 2005). Therefore, measures of oxygen are most useful when taken frequently over several days. The pattern of oxygen production and depletion can then be associated with nutrients more comprehensively than single point measures.

Chl-a concentration in samples scraped from substrates or in the water column is an estimate of algal biomass. It was assumed that greater amounts of chl-a would be associated with greater nutrient availability (Biggs 2000, Chetalat et al. 1999, Dodds et al. 2002) and greater productivity (measured as DO dynamics).

Benthic macroinvertebrates interact directly with DO and periphyton. The effects are through pathways of respiration (oxygen supply), habitat character and food availability. These variables are affected by nutrients, so macroinvertebrates are indirectly related to nutrients. If direct nutrient – macroinvertebrate response pathways exist, they are not well defined.

Macroinvertebrates graze and inhabit periphyton communities. Some grazers prefer certain types of periphyton (Calow and Calow1975, Lodge 1991). Excessive periphyton can degrade macroinvertebrate habitat for those organisms that require substrates with sparse algal growth (Downes et al. 2000). Therefore, the indirect effects of nutrients on benthic macroinvertebrates, through periphyton, can cause varied responses in the macroinvertebrate community. These interactions can occur in both directions – with macroinvertebrates effecting periphyton through selectively grazing, sometimes to a degree that affects nutrient uptake from the water column (Wallace and Webster 1996). Benthic macroinvertebrates are responsive to many stressors other than nutrients, and the possible confounding effects should be factored out, when they are recognizable. In a study of stream conditions in the Ozark Highlands Ecoregion in Arkansas, biotic indices for three biotic assemblages were negatively correlated to nutrient concentrations (Justus et al. 2010). The algal index had a higher correlation (rho = 0.89) than did the macroinvertebrate and fish indices (rho = 0.63 and 0.58, respectively). This suggests that the more direct effects with few indirect factors were reflected in the stronger correlations.

There are many modifying factors in a stream ecosystem. For example, phosphorus and light interact to effect algal growth (Hill and Fanta 2008, Hill et al. 2009, Mulholland 2001). Under controlled conditions it takes very little P to maximize algal growth given high light. However, the relationship may not be observed if algal production is limited by other factors such as bottom substrate, turbidity, canopy cover, hydrology, or depth. The fundamental relationships might be observed in the datasets, or the modifying and co-varying factors might mask or exacerbate the expected relationships. The modifying factors were used to help in natural classification to characterize multivariate relationships.

Conceptual models were developed to represent known relationships between changes in TN and TP concentrations, biological effects, and attainment of designated uses. Conceptual nutrient models are well established and were not reconstructed for this study. Instead, the conceptual model published by EPA (2010) was used as a standard that is applicable in New Mexico streams (Figure 1). The conceptual model shows intricate pathways of effects. It illustrates interactions that might be effective though our analytical data set is insufficient to account for them. Analyses that compared indirect elements in the conceptual model (e.g., relating nutrient concentrations to macroinvertebrate responses) relied on the validity of the intermediate linkages. The conceptual model provides a means of communicating the current state of knowledge regarding the effects of TN and TP in aquatic systems and is an important tool for guiding causal analyses.

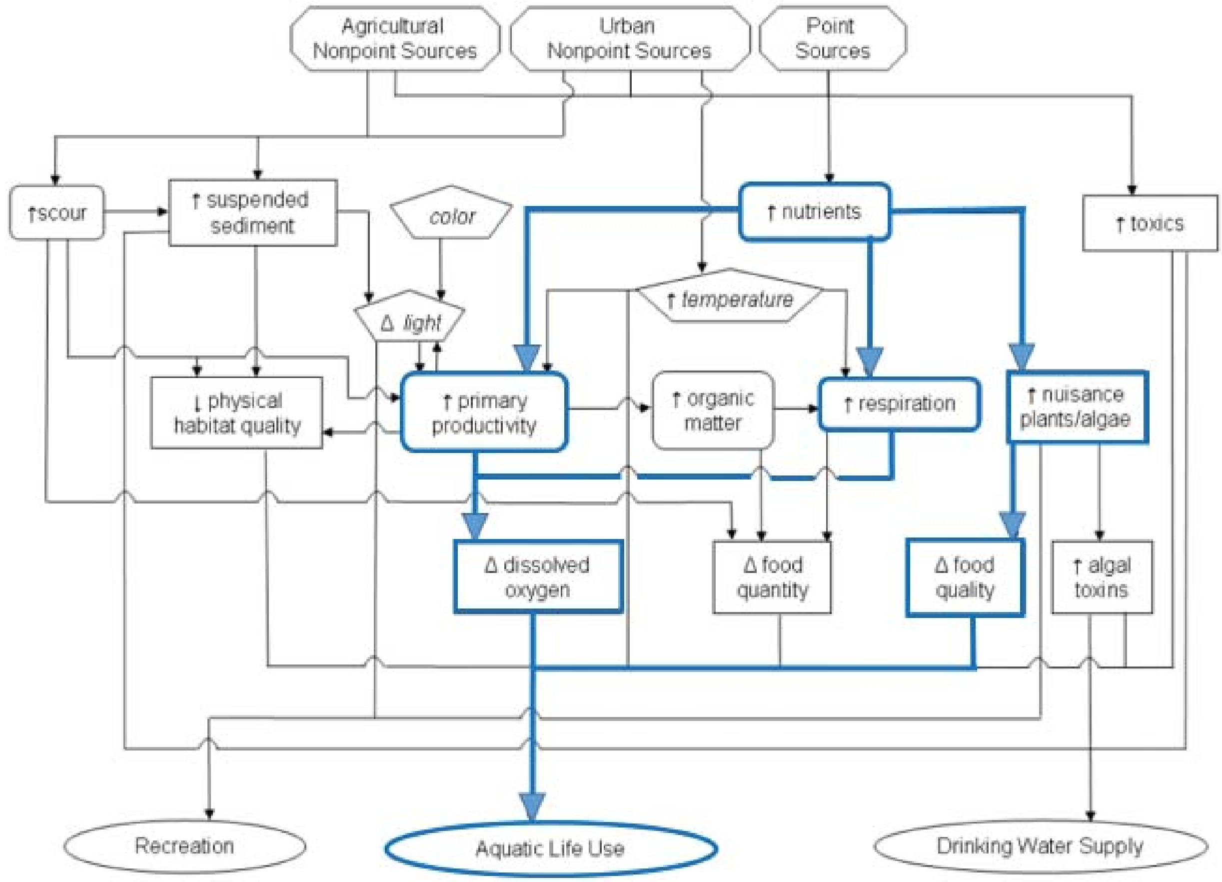


Figure 1. Conceptual diagram linking sources of human disturbance with designated uses through pathways that include nutrients (from USEPA 2010).

The linkages and relationships explored through stressor-response analysis are listed below. These relationships were analyzed as one-to-one stressor-response relationships, relationships with modifying factors, and as indirect relationships (e.g., nutrient-macroinvertebrate responses without intermediate links).

1. ↑ Nutrients = ↑ Chlorophyll a
2. ↑ Chlorophyll a = Δ DO dynamics
3. ↑ Nutrients = Δ Diatom Metrics
4. Δ DO dynamics = Δ Macroinvertebrate Metrics

These relationships were explored in stressor-response analysis for the aquatic life endpoints: diatoms and macroinvertebrates. The indirect relationships between nutrients – DO and nutrients – macroinvertebrate metrics were explored in addition to the direct nutrient – chl-a – DO – macroinvertebrate relationships. The indirect relationships were explored because data were lacking in all of the data types at all sites. The subsets of sites with chl-a and diel DO data are relatively small. Weak relationships in these small data sets might be due to sample size instead of a true lack of response. There are almost no sites with data from all data types collected in the same sampling season.

1.4 General Approach to Developing Thresholds

The analytical methods consisted of two major categories: indicator value distributions emphasizing reference conditions and stressor-response relationships. The reference condition approach includes identification of minimally disturbed sites, classification of the sites, and description of the reference condition based on characteristics in those sites in each class (USEPA 1998, 2000a, Barbour et al. 1999, Stoddard et al. 2006). In this approach, candidate thresholds were derived from distributions of nutrient concentrations in reference waterbodies, in this case, wadeable perennial streams.

Stressor-response approaches refer to analytical techniques that derive candidate thresholds by exploring the relationships between response variables and nutrient concentrations. Typical response variables in the context of nutrient threshold development include biomass and assemblage metrics (e.g., percent nutrient sensitive diatoms) and aquatic life use indicators or biocriteria indicators (e.g., trophic state indices, algal multi-metric indices, or invertebrate multi-metric indices). The value of these indicators is their association with aquatic life use designations. In New Mexico, assessment thresholds for biological indicators are established for only small, high elevation streams (Jacobi et al. 2006). This benthic macroinvertebrate index is used for assessment in this class of streams (NMED/SWQB 2013). This would provide a way to connect nutrient concentrations directly to aquatic life use protection. However, the small high streams do not cover all streams types that require thresholds. Therefore, biological metrics that were shown or presumed to respond to stressors were analyzed. The metrics represent the relative integrity of the aquatic community at a site. Maintaining metric conditions that are similar to conditions observed in reference sites is an indicator that designated aquatic life uses have been protected.

In EPA's draft guidance on Empirical Approaches for Nutrient Criteria Derivation (USEPA 2009), several methods for evaluating stressor-response relationships were presented. The approaches implemented in this analysis were adopted to take advantage of available data and to produce robust results using a combination of well-established and exploratory analytical techniques. The focus of the analysis was on the major nutrients, nitrogen and phosphorus, as they relate to the available response measures, periphyton (diatoms), chl-a, dissolved oxygen, and benthic macroinvertebrates.

The analytical techniques recommended for relating stressors and responses included correlation analysis, regression interpolation, and change-point analysis. These techniques are introduced here and are explained in greater technical detail in the methods section. Each technique has advantages and limitations that lead to differential weighting of results from the nutrient threshold analyses.

1.4.1 Reference conditions and classification

Reference streams represent least disturbed and/or minimally disturbed conditions (Stoddard et al. 2006), share similar characteristics with the waterbodies for which thresholds are being derived, and ideally support designated uses (US EPA 2000a). Reference sites were identified and a disturbance gradient defined using objective criteria for the stressors and stressor sources measured at the site or sensed remotely and analyzed using Geographic Information System (GIS). While quantitative criteria are defensible and repeatable, subjective input from knowledgeable experts was also used to confirm final reference site selection to account for factors that were not captured in the GIS analysis.

Natural, background, or reference nutrient concentrations can be inferred from conditions observed in reference streams. These are referred to as the reference conditions (Barbour et al. 1999, Stoddard et al. 2006). Reference nutrient conditions are subject to unavoidable human activities (such as atmospheric deposition), availability of suitable reference sites, and adequate recognition of natural variability.

Thresholds for a particular variable (e.g., TN or TP) were derived by first describing statistical distributions from reference waterbodies and then selecting a representative value to define expectations for reference and for all other waterbodies in that class. USEPA's nutrient criteria guidance recommends the use of percentiles derived from the reference waterbody distributions, since these waterbodies represent an example of the biological integrity expected for a region (USEPA 2000). The reference condition approach requires confident definition and identification of reference waterbodies, accounting for natural variability within site types, and availability of sufficient data from these reference waterbodies to characterize the distributions of nutrient variables.

Nutrient concentrations in New Mexico streams with minimal human disturbance were expected to have relatively low variability within homogenous landscape types. Site classification is the process by which natural gradients among sites are examined to identify sites with similar nutrient conditions in the absence of human disturbance. The purpose of classification is to minimize within-class natural variability of indicators so that anthropogenic disturbance can be recognized with less background noise (Barbour et al. 1999, Hughes et al. 1995). Potential site classification variables, nutrient indicators, and biological variables were analyzed simultaneously to identify patterns of covariance that suggested how nutrient conditions would be classified according to environmental and biological characteristics. The level III and IV ecoregions of New Mexico (Griffith et al. 2006) were considered as potential classification groupings, as indicated by previous analyses that distinguished five site classes (NMED/SWQB 2011). The grouping of the level IV ecoregions into mountains, foothills, and xeric as used in sediment assessments was also examined. In addition, NMED recognized coldwater, warm-water, and transitional (cool-water) designated uses, which are considered as another layer of site classification.

1.4.2 Frequency Distributions

Distributions of nutrient concentrations provided a baseline description of nutrient conditions throughout New Mexico. The distribution percentiles in different subsets of the data were used to describe general nutrient conditions by nutrient, site class, or disturbance status of the sites. These standards (percentiles of distributions in site categories) have long been established (USEPA 1998, USEPA 2000a, Barbour et al. 1999) and are now accepted as practical guidelines for describing reference expectations. This approach was also used in developing nutrient guidelines in New Mexico (NMED/SWQB 2011).

The 75th percentile of reference sites within site classes is a suggested value used for deriving potential thresholds for nutrients. At this level, a lower value passes the threshold and indicates that the observation is similar to the lower 75% of nutrient concentrations. If confidence in reference sites and site classes is high, less error in the reference values should be expected and a higher percentile of the distribution (such as the 90th) could be used as a guideline for establishing thresholds. If confidence is low, as when the best available reference sites are selected for a region that has a generally high intensity of disturbance, then more error in the reference sites can be expected and a lower percentile (e.g., the 50th) can be selected. An observation in the site class with low confidence in the reference conditions will only pass if it is similar to the best 50% of reference values.

The reference site selection process and the distribution of nutrient values observed in the reference sites were expected to have variability and error. Errors might be attributed to unmeasured stressors causing an incorrect reference designation, misclassified sites, unrecognized site classes in which natural conditions are unusual, or faulty measurements or recording of the nutrient endpoints.

1.4.3 Correlation and interactions

The Spearman's rho correlation coefficient and bi-plots of nutrient concentrations and response metrics were used to identify potential relationships between responses and nutrient variables. Correlation analyses identify the apparent linkages between biological condition and environmental variables. Bi-plots were examined to determine if the correlations reflected a feasible relationship.

Buffering or modifying co-variates were examined using multiple regression and recursive partitioning (also known as classification and regression trees (CART)). Variation of the response variable was explained along the gradient of nutrient concentrations and in relation to the covariates. If major covariates or modifying factors were recognized, attempts were made to account for them while estimating the direct effects in the major nutrient – biological pathways.

1.4.4 Regression Interpolation

When a clear linear relationship is evident between a nutrient concentration and a response variable with an existing threshold, then a nutrient concentration can be associated with the response threshold through intersection with the linear regression. For example in a bi-plot of the New Mexico macroinvertebrate stream condition index (NMMSCI) in high elevation streams with smaller watersheds, the value of TN that corresponds to the NMMSCI threshold value is determined as are ranges of TN values associated with error in the regression. When no response thresholds were established, the 25th or 75th quartile of the response metrics in reference sites was used to represent a protective threshold.

In this approach as in all others, variability that can be attributed to factors other than those being analyzed must be accounted for or otherwise recognized. To continue with the example above, relationships between nutrient concentrations and macroinvertebrate index values must include only sites for which the index is valid (high small streams), using consistent or comparable sampling methods, and factoring out known sources of variability (when feasible). When the modifying factors are not accounted for, then the linear regression could be driven by factors unrelated to nutrient.

1.4.5 Change-point Analysis

The change-point is the point along an environmental gradient (nutrient concentrations) at which there is a high degree of change in the response variable (chl-a, diatom, or macroinvertebrate metrics). The nonparametric deviance reduction method for identifying change-points (Qian et al. 2003, King and Richardson 2003) works well when the response is stepped, or drastically changing at a recognizable point along nutrient concentration gradient. With this method, the data are divided into two groups, above and below a potential nutrient threshold, where each group is internally similar and the difference among groups is high. This technique is similar to regression tree models, which are used to generate predictive models of response variables for one or more predictors. Using this comparison, the change-point is the first split of a tree model with a single predictor variable (i.e., nutrient concentration).

One caveat of the change-point analysis is that a change-point may be identified, and even determined to be statistically significant, when the change-point value is actually only an artifact of the analysis and not an indication of a change in system properties (Qian and Cuffney 2012, Daily et al. 2012). The methods can find change-points, even in datasets with nearly straight line relationships between X and Y. It has been well established that nutrient concentrations limit algal growth as well as species composition. Therefore, it is reasonable to believe an ecological threshold does exist between certain periphyton metrics and nutrient concentrations. We qualified the changepoint results using three assessment measures: confidence in the changepoint, coincidence of the changepoint with a break or extreme slope of the local (LOWESS) regression, and indications of a limiting relationship in a quantile regression.

1.4.6 Synthesis of Multiple Thresholds

The strength of an analysis with multiple approaches and response endpoints comes from the multiple lines of evidence. They can be used to show central tendencies and ranges in potential thresholds. The central tendency of potential thresholds shows corroborated evidence, which can give greater confidence in a selected threshold. However, values that deviate from the central tendancy may be selected if the regulating agency has a reason to adjust the level of protection. This decision may be based on confidence in an individual analytical technique or on corroborating evidence from the scientific literature.

The reference percentile selected, the stressor-response results considered, and the potential thresholds that result, must conform to the Clean Water Act (CWA) regulatory goals. The intended level of protection for designated uses should be clearly stated and used to justify the proposed thresholds. The selection of a percentile of the reference condition as a threshold is defensible when it is supported by clear reasoning and corroborating analyses.

NMED will select nutrient thresholds from the range of possible thresholds. Their selection needs to be communicated and justified transparently. To help with this, the potential thresholds from the multiple analyses and the scientific literature were tabulated for review. Each potential threshold was listed with any caveats, error, and uncertainty associated with the method or the underlying data. In addition, interpretations of the protectiveness of certain thresholds were described in terms of the designated uses.

2.0 Data Description

Data were collected from three sources: NMED monitoring programs, the National Rivers and Streams Assessment (NRSA) and the Wadeable Streams Assessment program (WSA) (Table 5). The NRSA and WSA data were collected under USEPA programs in which NMED participated. In addition, a GIS analysis of sites and their catchments was conducted to characterize sites.

All data were compiled in a single relational database (Microsoft Access), though data from each source were maintained in separate database tables. Data compilation and analysis were conducted following the Quality Assurance Project Plan (QAPP) (Tetra Tech 2011a, 2012). The QAPP primarily addresses procedures for working with secondary data. Procedures for data compilation and analysis were followed to minimize errors in transferring data from original sources into analytical databases. Although data were expected to be error-free in the original form, exploratory analysis techniques (e.g., correlation analysis, scatter plots, histograms, etc.) were used to identify potential outliers and other data quality issues.

Table 5. Data summary by source.

NMED:	883 valid sites in NM with water chemistry (targeted sampling design) Multiple samples per site (approximately 7352 samples) Years 1990 - 2012 Chemistry, site & habitat characteristics (partial data depending on site and visit) Benthic macroinvertebrate samples in 202 sites (440 samples) Periphyton (diatoms) in 212 sites Benthic chl-a in 146 sites Dissolved oxygen diel data in 175 sites
NRSA:	88 sites, each with a single visit (probabilistic sampling design) Years 2008 - 2009 44 sites in NM, others within 50-150 miles of NM Chemistry, benthic & sestonic chl-a, periphyton, site & habitat characteristics
WSA:	56 sites, each with a single visit (probabilistic sampling design) Years 2000 - 2004 10 sites in NM, others within 50-150 miles of NM Chemistry, benthic macroinvertebrates, site & habitat characteristics

The NMED data were more numerous than NRSA or WSA data. They were collected for various projects over time. The NMED data included four data types: nutrients and other water quality analytes, periphyton (benthic chl-a and diatoms), macroinvertebrates, and diel dissolved oxygen. The NMED monitored four primary water quality variables in New Mexico streams (TN, TP, benthic chl-a, and turbidity) plus a number of secondary variables including DO concentration, DO percent saturation, and pH (NMED/SWQB 2008). The NMED nutrient records were most numerous and complete, while other types of data were relatively sparse and only used in analysis when associated with nutrient information.

NRSA and WSA data were obtained from the USEPA Region 6 and the Office of Research and Development in Corvallis, OR. They included information regarding water chemistry, physical habitat, and biological assemblages. For the NRSA, information on benthic macroinvertebrates, periphyton, and both benthic and sestonic chl-a were available. For the WSA data, benthic macroinvertebrates were the only biological data available. Data for NRSA were collected in accordance with the Field Operations Manual (USEPA 2007). For WSA, data were collected using methods similar to those used in the NRSA, following procedures outlined by the USEPA (2004) and Peck and others (2006).

2.1 Sites

The study area included the state boundaries of New Mexico as well as regions in adjacent states that were within 50 miles of New Mexico and in a level 3 ecoregion that also existed within New Mexico. To the north and east of New Mexico, sites from further away (up to 150 miles) were considered because a lack of sites in the drier ecoregions was anticipated. NMED data were collected as part of the nutrient thresholds development projects and for other water quality surveys. Some of these data were targeted to focus on reference reaches for stream classification and for identifying threshold values for nutrients, algal biomass, and secondary variables (NMED/SWQB 2008). The streams were selected to span at least five ecoregions throughout the state. Monitoring focused primarily on a critical low flow index period from August to November. At some sites, the monitoring plan allowed examination of seasonal and annual variability and trends.

A uniform Station ID was assigned to each site based on the identifier used in the original surveys. In some cases, especially within the NMED data set, some sites were re-visited, but given a slightly different Station ID. Such cases were identified through GIS and NMED site list reviews. Station IDs were adjusted to reflect co-occurring sites. The Station ID was used to link information among databases. Each Station ID was associated with latitude/longitude coordinates that were used in GIS analyses.

Site characteristics

Site characteristics were either observed or measured in the field, or they were remotely sensed and derived from GIS analysis. The observed or measured data were recorded during site visits and included physical habitat assessments, channel dimensions, slope, canopy cover, riparian vegetation, riparian integrity, substrate characterizations, flow, and more. Each data source (NMED, NRSA, and WSA) included somewhat different variables for the observed and measured site characteristics. For NMED, habitat and flow variables were not collected at each site, but were often associated with benthic macroinvertebrate samples.

The remotely sensed data were derived from GIS analysis. These data include information on the setting of the sampling site and surrounding areas, such as ecoregion, land use types and intensity, roads and road crossings, population density, watershed area, and more (Table 6). The GIS variables were either for site classification or for disturbance evaluations. Analyzing every site in the NMED data set was not possible because we needed to be efficient with the GIS analysis task. The GIS analysis was conducted on a subset of sites prioritized based on data completeness and potential reference site status. There were several NMED sites that were not included in GIS analysis, including those with only nutrient data that showed no preliminary indication that they were reference quality. High priority sites were identified using the following selection criteria.

1. Sites with nutrient data and comparable response data had high priority. Sites that had periphyton (diatoms), chl-a, diel dissolved oxygen, or benthic macroinvertebrate data were valuable for detecting responses to nutrient conditions.

2. Potential reference (least disturbed) sites were high priority. These were used to characterize background nutrient conditions in relatively undisturbed sites.

A thorough GIS analysis was performed based on delineated catchments upstream of 660 sites. The GIS information is consistent throughout the project area, including sites sampled by multiple data collection programs. The information summarized from the GIS analysis includes land use, human activities, and environmental characteristics that are appropriate for reference site designation and site classification. Additional details are provided in Appendix B.

Table 6. Variables used in GIS analysis.

Variable	Description
Point Values	
Stream Slope	NHD Plus join with flowline attributes table
Stream Order	NHD Plus join with NHDFlowlineVAA table
Elevation (cm)	NHD Plus DEM files
Designated Use	RAD 305b Assessed Segments joined with ATTAINS data
Precipitation	PRISM
Temperature	PRISM
Level 3 and 4 Ecoregions	EPA Ecoregions
Geology	USGS Integrated Geological Map
Watershed Values	
Road density	Attila tool and TIGER 2000 files
Number of road/stream crossings	ARCGIS tools
Land Slope	ARCGIS Spatial Analysis Slope tool
Land Use and Cover	Attila tool and NLCD 2006 data
Canopy Density	Attila tool and NLCD 2001 Canopy data

2.3 Nutrients and Water Quality

The NMED nutrient database included more than 7,000 records of nutrient concentrations at 883 stream sites (including sites not GIS analyzed). Samples used in this project were collected from 1990 to 2012. The nutrients recorded were related to nitrogen (ammonia, NO_3NO_2, and TKN), phosphorus (orthophosphate and total phosphorus) and ancillary analytes (pH, specific conductance, temperature, turbidity, and dissolved oxygen). Total nitrogen was calculated as NO_3NO_2 + TKN. Measures of total phosphorus were roughly 24x more common than measures of orthophosphate, and therefore orthophosphate was not analyzed.

NMED collected and processed samples in accordance with methods documented in an EPA approved Quality Assurance Project Plan (QAPP) and associated Standard Operating Procedures (SOP). The QA/QC procedures in the QAPP included collection and analysis of replicates for 10% of water samples, adherence to calibration methods, and taxonomic verification of a subset of periphyton and benthic macroinvertebrate samples. Also included was a thorough QA review of all site and analytical data, including flagging of all parameters that were outside of the control limits.

Nutrient data from the NRSA included TN and TP as well as nitrate, nitrite, and ammonium. These data plus information on the ancillary variables pH, specific conductance, temperature, turbidity, dissolved oxygen, physical habitat, benthic macroinvertebrates, and diatoms were complete for all 88 sites, half of which were in New Mexico. WSA nutrient data included TN and TP as well as nitrate and ammonium. Data for water chemistry including ancillary analytes, physical habitat, and benthic macroinvertebrates were complete for 56 sites, 10 of which were in New Mexico.

Diel Dissolved Oxygen

Diel dissolved oxygen data were collected by NMED in stream sites throughout New Mexico between June and October (mostly August and September) from 2001 to 2012. These data were collected along with pH, specific conductance, temperature, and turbidity using multi-parameter, continuous recording sondes with recording periods of at least 48 hours and recording intervals ranging from 15 – 60 minutes. Multimeters were generally placed in deeper shaded pools of moving waters where they were stable and not conspicuous. The data from approximately 200 spreadsheets were combined in a single data set so that metrics could be calculated efficiently. Data were checked for errors and data points or whole records were revised or eliminated if they were perceived to be inconsistent. Some of the turbidity data showed erratic patterns and continuous turbidity was not analyzed. Errors that typically occur with sonde data relate to records before and after the sonde is placed in the water or in association with drifting calibration. For dissolved oxygen, drift in calibration was only suspected in early years, when probes with membranes were used. They were replaced with optical sensors over time. After QC, statistics on 175 diel DO records were calculated. Diel DO statistics were related to nutrients in 133 sites, one sample per site. Diel DO and chl-a measurements coincided in 64 sites. Sestonic chl-a measures were not taken at sites with diel DO data.

The metrics that were calculated included, but are not limited to, overall minimum DO, maximum daily fluctuations, and standard distribution statistics. NMED provided metrics on the maximum productivity and respiration in each data set based on 2, 3, and 4 hour intervals and the 4-hour interval was used in analyses. Distribution metrics were also calculated for DO percent saturation data. In addition, system metabolism was calculated as gross primary production (GPP) and ecosystem respiration (ER), which accounted for temperature, elevation, and estimates or derivations of barometric pressure, nighttime regression, and light exposure. The calculations were carried out in R software using code provided by Dr. Robert Hall (Department of Zoology and Physiology, University of Wyoming, Laramie, WY).

Data Reduction

All of the data should be considered for analysis, but some analyses are better suited to specific data types or summaries. There were four issues we addressed in reducing data for analysis.

1. Summarizing nutrient data for analysis
2. Identifying and eliminating outliers
3. Establishing estimated values for censored (non-detect) data
4. Limiting data to address seasonal variability

Summarizing nutrient data for analysis

In the NMED dataset, there were multiple samples collected at the same site over time. The median, geometric mean, or maximum of nutrient values per site were considered for describing value distributions and for site classification. Median site values were used to summarize site nutrient conditions.

For stressor-response analyses, the chemistry and response samples were limited to those collected within one month of each other. If multiple chemistry samples were collected during that period, the average value was used. Only a single stressor-response dataset was used per site so that sites with multiple response records over time would not bias patterns derived from multiple sites, most of which had only a single response record. Logarithmic transformations (base 10) were used to reduce skewness in nutrient concentrations and for other variables as needed.

Identifying and eliminating outliers

Outlier values in the database are expected to be associated with data entry errors, field and laboratory analytical errors, and anomalies in sampling conditions, such as elevated storm flows and runoff from fire damaged landscapes. Errors were not expected in large frequency because QC procedures were in place when the data were generated. Nevertheless, the database was searched for nutrient and other values that were unusual, inexplicable, or associated with anomalous sampling conditions. Approximately 150 of 10,000 NMED nutrient records were removed from analytical data sets due to high outlier values, most of which were associated with storm flows and fire runoff. The NRSA and WSA data had qualifier data that indicated QC review and did not have apparent outliers. Details of the outlier analysis are in Appendix C.

Establishing estimated values for censored (non-detect) data

Several NMED data points were designated as “non-detect”, having concentrations less than the sensitivity of the analytical equipment. For example, 17% of TKN values were flagged as non-detect in the NMED data. The current default for these values has been substitution of ½ of a standardized detection limit for all samples with values below that standard or marked as below detection at a higher reported detection limit. Alternative treatments were considered for the censored TKN, NO_3NO_2, and TP data, including elimination and adjustments using Kaplan-Meier (KM), regression on order statistics (ROS), and maximum likelihood estimation (ML). Based on a limited analysis (Appendix D) and review of similar data sets, half detection substitutions were used for all analyses. The half detection limit values were 0.015 mg/L for TP, 0.05 mg/L for NO_3NO_2, and 0.05 mg/L for TKN. In calculating TN from NO_3NO_2 + TKN, the value resulting from two non-detects would be 0.1 mg/L. The NRSA and WSA data had less than 5 values below the laboratory reporting limits of 0.02 mg/L TN and 0.004 mg/L TP. These low values were re-established at the reporting limit for analysis.

Limiting data to address seasonal variability

The NMED nutrient data were mostly collected in the spring, summer, and fall seasons. Samples were much less common in winter months (December, January, and February). Through the years, seasonal effects on nutrient concentrations were expected due to variable discharge or fertilization patterns, changes in light intensity, and variable rates of runoff and plant uptake. Therefore, patterns that might affect threshold development or application of thresholds were reviewed in the NMED nutrient concentrations over seasons as well as trends from individual sites with multiple records over time. The conclusion of the analysis (detailed in Appendix E) was to remove samples collected in the winter months (December, January, and February) from the general analysis. All of the data were accepted from the NRSA and WSA programs because they were collected within a narrowly defined index period (May through September). The numbers of samples and of years per site ranged up to 65 and 10, respectively, in the Rio Hondo site 28RHondo014.8 (Appendix F).

2.4 Response Measures

The response data were analyzed as metrics of each assemblage. The biological metrics are usually limited to those that are basic and familiar summary metrics or are known indicators of stress. Since there are more ways to measure an assemblage than there are meaningful ways to interpret several metrics, the number of metrics were limited through a preliminary screening process that discerned familiar, proven, precise, or sensitive metrics. Metrics that had high measurement error or were unresponsive to stress were not used for stressor-response analyses.

Chlorophyll a

Of the NMED wadeable stream sites with nutrient data, 174 also had benthic chl-a data (including 35 with benthic macroinvertebrate data as well). These samples were collected between 2004 and 2011 in the months of August to November. Chl-a data were also collected for 50 NRSA sites, including both benthic and water column measures.

For NMED samples, chl-a samples were extracted with 90% ethanol and analyzed with a spectrophotometer using a modified Standard Method for the Examination of Water and Wastewater, American Public Health Association, Method 446.0 (APHA 2012). The absorbance correction for ethanol (28.66 = absorbance correction for chlorophyll in ethanol) was substituted for the acetone correction of 26.7. Extraction was done according to methods in Biggs and Kilroy (2000), (extraction in ethanol boiled for 5 minutes and soaked for 12-18 hours).

Periphyton Data

Periphyton data in and around New Mexico were collected by NMED and the NRSA. Through the NMED, roughly 212 diatom samples were collected from 2002 to 2008 mostly in the fall sampling season (August - November). Soft algae samples were collected from 133 sites. Samples were collected using a targeted richest habitat sampling method (NMED 2014), which include scraping delimited areas from 5 – 9 cobbles, woody debris, or soft substrates within transects of the stream, preservation with Lugol's or formalin solution, and identification of 500-600 valves in the laboratory. Periphyton data from 69 NRSA sites in and around New Mexico were added to a single periphyton database. The NRSA periphyton samples include both diatoms and soft algae. Potential bias that might be introduced by different sampling protocols was investigated by comparing metric distributions.

For the NMED and the NRSA diatom data, metrics were calculated in a relational database. Approximately 68 diatom metrics were calculated including metrics and taxa attributes described by Porter et al. (2008), Stevenson et al. (2008), Kelly and Whitten pollution tolerance index (1995), van Dam et al. metrics (1994), and periphyton indices developed by Potapova and Charles (2007). Eight responsive metrics were selected for continued analysis in stressor-response analyses

Benthic Macroinvertebrates

SWQB monitored benthic macroinvertebrate community composition at targeted sites. Macroinvertebrate and chemistry samples collected within 30 days of each other were identified in 202 NMED sites. One benthic macroinvertebrate sample was compared to average site chemistry from samples collected within 30 days of the benthic sample. If multiple benthic methods were used on a single date, a preferred method was selected, with preferences as follows: Reachwide > Kick > Targeted Riffle > other.

The NMED macroinvertebrate samples were collected using six different methods, including reachwide, targeted riffle, kicknet, surber, Canton Hess and Jacobi Hess. Biomonitoring samples were collected in accordance with the EPA Rapid Bioassessment Protocol (RBP) (Barbour et al. 1999), the NMED Standard Operating Procedures (SOP) (NMED/SWQB 2005, 2012), and/or modified EPA EMAP macroinvertebrate sampling method (Peck et al. 2006). Opportunities to aggregate samples collected by different methods were explored and samples from multiple methods were pooled when the results of each method overlap in stressor-response bi-plots. Separate analyses were conducted for methods that could not be aggregated because of non-overlapping data points in the bi-plots. NRSA and WSA benthic data were collected with consistent reachwide or targeted riffle methods (Peck et al. 2006) and were summarized as metrics in spreadsheet format. In the WSA and NRSA datasets, 56 and 40 benthic samples (respectively) matched the chemistry samples.

NMED benthic samples were the basis for calculation of 63 metrics in categories of taxa richness, composition, pollution tolerance, feeding groups, and habit. These metrics were used in assessments or were expected to be responsive to stresses in New Mexico streams. Ten benthic macroinvertebrate metrics with consistent and strong correlations were identified. The high-small multi-metric macroinvertebrate condition index (HSMMCI, Jacobi et al. 2006) is an assessment index that NMED uses in sites with elevations >7500 feet and catchments <200mi^2.

3.0 Methods

The general approach to developing nutrient thresholds includes frequency distributions of the data and stressor-response analysis. For frequency distributions, a disturbance gradient was developed, sites were assigned to reference categories, and site classes were established to reduce natural variability. The stressor-response analysis includes regression interpolation and change-point analysis techniques. Diel DO measures were analyzed for thresholds related to reference conditions, nutrient concentrations and macroinvertebrate responses.

3.1 Reference Sites and Classification

Reference Site Identification

Reference sites were needed in the New Mexico nutrient analysis for characterizing the nutrient conditions in the absence of substantial disturbance. This allowed exploration of natural variation in nutrient concentrations across the study area and derivation of potential nutrient thresholds from distributions of nutrient values in the relatively undisturbed sites. Stream classification and reference site designations hinged upon each other to characterize nutrient conditions relative to both natural and disturbance gradients.

Reference stream sites have been identified in and around New Mexico for multiple purposes, including biological index development (Jacobi et al. 2006, Paul 2008), sediment threshold estimation (Jessup et al. 2010), and the national stream surveys. The designations established for each purpose were adopted to create a list of potential reference sites for this project. Reference designations established for the national surveys used nutrient measures as criteria for ranking disturbance levels (Kaufmann et al. 2012). This was inappropriate for our analysis and the national surveys reference designation (and all others) were reevaluated using GIS analysis to confirm the designations.

For the first cut of reference site designations, a thorough GIS analysis was performed based on delineated catchments upstream of each site. The GIS information was derived for 662 sites in the project area, including sites sampled by multiple data collection programs. The reference designations established in other studies and NMED staff review of empirically derived designations were used for confirmation. Because the NMED staff are familiar with site conditions that may not be reflected in the GIS data, they reviewed the reference designations indicated through empirical analyses and were able to change designations based on knowledge of the sites. For example GIS coverages do not reflect the intensity of grazing. Sites with contradictory indications of reference status from the multiple techniques were relegated to non-reference or “Other” categories.

Land use coverage and human activity in the catchments (Table 7) were examined for appropriate thresholds to indicate disturbance or lack of it. Development and agricultural land uses indicate catchment scale intensity of disturbance. Both pasture and crops were considered agricultural uses. Forest, water, and wetland are usually undisputable natural land covers. However, the “natural-ness” of scrub/shrub, grassland, and barren coverages are uncertain because they could be due to human activities or natural environmental factors, especially in more arid areas. Therefore, the known disturbances were emphasized. Road density and the density of road-stream crossings were used as a surrogate for intensity of human activity in the watershed.

Table 7. Variables used as reference site criteria (see Appendix B for additional details).

Variable	Description	Source
Urban Index	% low, medium and high intensity development in the catchment	CDL_NLCD (2010)
Agricultural index	% pasture/hay and cultivated crops in the catchment	CDL_NLCD (2010)
Road Density	Length of roads per area (mi/mi^2) in the catchment	Roads (Census Streets, 2010)
Road Crossing Density	Count of road-stream crossings per area ($\#/mi^2$) in the catchment	Road/Stream Crossings
Dam Density	Count of dams per mi^2 in the catchment	Dams and Diversions (NHDPlus V2)
NPDES Density	Count of NPDES discharges per mi^2 in the catchment	NPDES Permits
Superfund Density	Count of Superfund sites per mi^2 in the catchment	Superfund Sites
Mine Density	Count of producing mine sites per mi^2 in the catchment	Mines (MRDS, Producers and Past Producers)
Dam Distance	Minimum distance of a dam to the sampling coordinates (as the crow flies)	Dams and Diversions (NHDPlus V2)
NPDES Distance	Minimum distance of a NPDES discharge to the sampling coordinates (as the crow flies)	NPDES Permits
Superfund Distance	Minimum distance of a Superfund site to the sampling coordinates (as the crow flies)	Superfund Sites
Mine Distance	Minimum distance of a producing mine site to the sampling coordinates (as the crow flies)	Mines (MRDS, Producers and Past Producers)

The known human activities in the catchment (dams, NPDES permits, Superfund sites, and mines) were used to qualify reference sites. Information on these activities were available as counts in the catchment, densities (counts/catchment area), and distance to the sampling site. Because sites have variable size, densities were used to standardize the activities per unit of land area. The distance between the activity and the site weights the probable effects of the activities at the sites. Activities greater than 5 miles distant were assumed to have less effect on site conditions than those within 1 mile of the site. Dams and NPDES permits were always on the stream network. The mines in the database were limited to those that were current or past producers. Mine locations that never went into production were excluded. Site specific stressors (in-stream measures of water quality and habitat) were excluded as reference site criteria because they could introduce circular logic in the threshold development process and they often represented only instantaneous conditions (USEPA 2013).

For each of the reference criteria variables from the GIS analysis, thresholds were established for five disturbance categories from reference to extremely stressed sites (Table 8). The thresholds were derived from distribution statistics for each criterion in all sites. The percentiles were used as guidelines for establishing thresholds, but subjective adjustments were made to arrive at feasible values and adequate numbers of sites in each disturbance category. Five disturbance categories were defined from best to worst conditions: Reference, Near-Reference, Other, Stressed, and Extremely Stressed. Reference sites are more important to identify than stressed sites, but stressed sites allowed us to recognize a full scale of disturbance gradients. To receive reference status, a site must not fail any of the stressed criteria and must pass at least 7 of the 8 reference criteria. Near-Reference sites did not fail any of the stressed criteria and passed at least 7 of the 8 reference Near-Reference criteria. Near-reference sites were used in analyses when reference sites alone were too sparse regionally or in a specific analysis. The inclusion of near-reference sites is declared for each analysis that follows. Stressed and Extremely Stressed sites failed at least 2 of the Stressed or Extremely Stressed criteria, respectively.

Table 8. Reference and stressed site criteria, based on distributions of values over all 660 sites.

Variable	Reference	Near Reference	Stressed	Extreme Stress
Urban Index (% cover)	0.01	0.02	1	2
Agricultural index (% cover)	0.1	0.5	4	5
Road Density (mi/mi^2)	1	1.4	3	5
Road Crossing Density (#/mi^2)	1	1.25	2	5
Dam Density (#/mi^2)	0	0.005	0.03	0.05
NPDES Density (#/mi^2)	0	0.01	0.1	0.2
Superfund Density (#/mi^2)	0	na	0.01	na
Mine Density (#/mi^2)	0.05	0.1	5	10
Dam Distance (mi)	na	na	1	0.5
NPDES Distance (mi)	na	na	1	0.5
Superfund Distance (mi)	na	na	2	1
Mine Distance (mi)	na	na	0.5	0.25

Reference sites identified through application of the GIS derived criteria were checked against previous reference designations where they existed. Sites with contradictory indications of reference status from the multiple techniques were relegated to the non-reference ("Other") category. The NMED staff was familiar with site conditions that were not reflected in the data (e.g. spills, fires, legacy effects) and they used this knowledge when reviewing the empirical disturbance categories for each site.

Site Classification

Natural gradients in the dataset that affect potential nutrient and biological response indicators were examined by first isolating those sites with minimal disturbance and then using appropriate statistical methods (e.g., principal components analysis, correlation analysis, and examination of bi-plots and distributions, etc.) to develop a stream classification scheme that captures the environmental variability for subsequent statistical analyses. Aggregate ecoregions used in the EMAP-West study (Stoddard et al., 2005)—Mountains, Plains, and Xeric—were considered as a starting point for stream classification, but were modified as needed and considered along with other categorical and continuous variables. The classification scheme developed for sediment assessments – Mountains, Foothills, and Xeric areas (Jessup et al. 2010) was also tested. Additional classification categories and variables were examined, including Level III and IV ecoregions (Griffith 2006), geology, latitude, longitude, stream order, elevation, drainage area, average land slope in the catchment, average annual precipitation, average annual temperature, width/depth ratio, entrenchment ratio, sinuosity, channel substrate, and stream slope.

The first step in site classification is defining the data frame, or population of streams from which data will be analyzed and to which resulting thresholds can be applied without extrapolation. The waterbodies of interest for this effort include perennial wadeable streams in New Mexico and in close proximity and similar ecoregions of neighboring states. It does not include ephemeral or intermittent streams, springs, and direct WWTP effluent. Also excluded are large rivers, that cannot be monitored effectively with methods developed for wadeable streams and generally have drainage areas greater than 2,300 square miles (NMED/SWQB 2011). The systems considered to be large rivers, and consequently <u>exempt</u> from this protocol, include:

1. San Juan River from the Navajo Nation to the Navajo Reservoir,
2. Rio Grande in New Mexico,
3. Pecos River from the Texas border to Sumner Reservoir,
4. Rio Chama from the Rio Grande to El Vado Reservoir (to due flow augementation from the San Juan/Chama project),
5. Canadian River from the Texas border to the Cimarron River, and
6. Gila River from the Arizona border to Mogollon Creek.

Principal components analysis (PCA)

Principal components analysis (PCA) was used as a tool for selecting site classification variables. The PCA was run in two configurations: only reference and near-reference sites and all sites. When only reference and near-reference sites were used, natural, nutrient, and stressor variables were included as determinants. When all sites are included, only natural variables were included. The advantage of using all sites is that regions with fewer reference sites will be represented. When included as supplemental variables (not influencing the organization of principal components), the stressor, biological, and nutrient variability can be compared to the principal axes. Nutrient-related axes that were correlated with biotic variables were examined to gain insight into potential scaling or classification variables that would minimize biological variability and thus focus biological responses on disturbances. Variables were transformed as needed to approximate normal distributions using logarithmic and Arcsine-Square Root transformations. Ecoregion designations and other classification variables were entered as binary code (true or false). For the PCA and other classification exercises, variables were as described in Table 9 and Appendix B.

Table 9. Classification variables.

Code	Description	Type
Latitude	Latitude	Continuous
Longitude	Longitude	Continuous
Elev_m	Elevation (m)	Continuous
DrAreaMi2	Drainage area (square miles) (log transformed)	Continuous
LndSlpAvgpct	Average land slope (%)	Continuous
PrecipAvg30	30 year average precipitation (mm)	Continuous
TempAvg30	30 year average air temperature (C)	Continuous
NMEDnutClass	NMED existing nutrient classes	Categorical
MFX	Mountain, Foothill, and Xeric classes	Categorical
Ecoreg3	Level 3 ecoregion	Categorical
GeolRockType1	Geologic rock type	Categorical

Correlation analysis was used to describe single factor relationships between nutrients and environmental variables in reference sites. In contrast to the PCA, the correlation analysis was always limited to reference sites to emphasize the effects of natural site conditions instead of disturbance levels. The non-parametric Spearman rank order correlation coefficient was used because it is less sensitive to skewed distributions.

The relationships suggested by PCA and correlations were examined in box plots and bi-plots. For example, the distribution of nutrient concentrations in reference sites of the existing site classes were plotted and examined for precision within classes and differences among classes. Distributions that showed high variability within a class indicated a need for more refined classification. Classes with similar interquartile ranges show possibilities for combining classes. If the box plots showed precise and distinct distributions, then confirmation of proposed classes was indicated. Bi-plots were used to show patterns of relationships between variables and to highlight tertiary attributes of the relationships such as reference status, ecoregion, or other covariants.

It is likely that multiple factors affect response variables, including stream shading, flow, substrate, turbidity, and others. The degree to which these other factors mask or accentuate responses between nutrients and response measures can sometimes be recognized and factored out. However, extensive analysis to recognize possible factors can suggest data parsing that is unreasonable for the sample sizes available for analysis. If the nutrient-periphyton analysis was reduced to one type of sample with similar characteristics (e.g., NRSA sites with >75% canopy, cobble dominated substrates, in cold water Southern Rockies streams), there would be few data points from which to find meaningful relationships. Therefore, while the distinct, categorical site classification does not account for all possible natural variability in nutrients conditions within each site class, nutrient observations could be adjusted based on regression residuals for the co-varying factors. This was further explored in the section on correlations and interactions.

Recursive Partitioning

Classification and Regression Tree (CART) models (also called recursive partitioning) account for variation in a dependent variable by progressively splitting samples into two bins that best partition the total variation among samples. This process forms a prediction tree based on a series of binary splits in the data. The first split occurs at the value of the predictor variable that most efficiently (as measured by the mean within-group standard deviation [SD]) partitions overall variation of the dependent variable into two groups. CART then partitions each of these two groups, if justified, into two smaller groups or nodes in the same manner, although the partitioning variable may differ. Trees were pruned to the complexity parameter value associated with the lowest cross-validation error. CART models were built with the R routine, ‘rpart’ (version 3.0.2; R Development Core Team, http://www.r-project.org/), for both TP and TN.

The CART analysis results in a cross validation error for each additional split of the model. When the cross validation error is minimal, the number of splits is maximized. Additional splits can overfit the data – where the model is very good at predicting within the training data, but is poor at predicting for new data. The splits can inform site classification – giving variables and thresholds that partition the data by nutrient levels. At the end of each branch of the tree, TN or TP values are predicted as the average for that group.

A random forest routine (R: randomForest) was conducted to find the most important variables in 500 runs of CART using random subsets of the data for each run. The variable importance is the relative frequency with which each variable entered into the models. Importance can be used to confirm the selection of variables in the predictive CART model. Random forests do not generate a predictive model. At first, only quantitative variables (Table 9) were used to predict splits relative to site average TP and TN (log transformed) in reference and near-reference sites.

Categorical variables were added to the models to determine whether existing classifications were as strong as the quantitative variables.

3.2 Frequency Distributions

Non-parametric quantiles were calculated from frequency distributions of nutrient concentrations in subsets defined by site classes and disturbance category. The distributions were based on median TN and TP concentrations in each site. NMED preferred to use the median values as the best representation of site conditions because the data were log normally distributed and the median was a better estimate of the central tendency of the data. Also, mean values can be biased by few extreme values. The frequency statistics included data from all data sources for nutrients. The reference and near-reference sites were combined after confirming that the nutrient distributions were similar. It was assumed that statistics from a larger data set would be more robust than those from reference sites alone.

The non-parametric median, 75^{th}, and 90^{th} quantiles of the reference site summary nutrient concentrations were reported. The 75^{th}, and 90^{th} quantiles were candidate thresholds below which values were characteristic of most reference sites. Confidence intervals (90%) were calculated for each quantile using 1001 bootstrap iterations. Analysis was conducted using R software.

3.3 Correlations and Interactions

In the correlation analyses, relationships were examined as outlined for the stressor-response approach, including the following linkages:

1. ↑ Nutrients = ↑ Chl-a
2. ↑ Chl-a = Δ DO dynamics
3. ↑ Nutrients = Δ Diatom Metrics
4. Δ DO dynamics = Δ Macroinvertebrate Metrics

Spearman non-parametric correlation was used to explore direct one-to-one relationships not only for these primary linkages, but also for the indirect nutrients – DO relationship and the covariates that were suspected to have strong direct or indirect effects. GIS-derived site characteristics were consistently available for all sites. These variables included ecoregion, land use types and intensity, roads and road crossings, population density, watershed area, and other variables (see Table 6). Site measurements of water quality and chemistry, physical habitat features, channel dimensions, slope, canopy cover, riparian vegetation, riparian integrity, substrate characterizations, flow, and more were measured at some sites, but not consistently in all sites. The commonly measured variables were included in analyses of covariates. Canopy cover and flow were two variables that were assumed to modify relationships between nutrients and algal growth, but that were not consistently measured with all samples.

The analysis resulted in correlation matrices for all data available or in subsets by site class and data source. Scatter plots were used to illustrate relationships with strong correlation coefficients to check that the relationships were not driven by a few extreme values. Scatter plots were also used to highlight tertiary variables and covariates by marking symbols in categories.

Forward stepwise multiple regression analysis was used to predict responses based on several variables. Multiple regression is a technique that prioritizes the most influential variables for the selected dependent variable. Multiple variables can enter into the regression equation, though usually with decreasing effectiveness and reduced significance after the first 2-3 variables.

As described for site classification, random forest analysis was used to find relative importance among sets of predictor variables for each dependent variable of interest. The output of the random forest analysis is a prioritized list of variable importance in the multiple models. This was another technique used to find strong modifying variables in the expected relationships.

The correlation analysis was also used to reduce our long list of macroinvertebrate and periphyton metrics. More metrics were calculated than could be meaningfully interpreted. Those metrics that showed relatively strong and consistent responses among the nutrient, DO and chl-a conditions were retained for further analysis.

3.4 Regression Interpolation

Simple linear regression provides an estimate of the linear relationship between a response variable and an explanatory variable (such as TN or TP). The regression results in coefficients for the intercept and slope of a straight line representing the relationship between the two variables. A stressor-response relationship estimated by linear regression predicts the value of the response variable, given a particular nutrient concentration. Hence, if the value of the response variable that supports the designated uses is known for a waterbody, the stressor response relationship can “translate” this response threshold to a numeric criterion value (USEPA2010). In our regression analysis, the simple linear regression was calculated between response variables (macroinvertebrate and diatom metrics, diel DO statistics, and chl-a) and nutrient stressors. TN and TP were log transformed. All of the samples were matched within a ±30 day sampling window.

The critical quartile of the reference distribution (25th or 75th) was determined for each response metric in each site class. The quartile values were used in lieu of response impairment thresholds for the biological assemblages, because such thresholds are not widely established in New Mexico. For metrics that decreased with increasing stress, the 25th quartile was used, assuming that higher metric values represent conditions similar to most reference sites within the site class.

Stressor values for each metric were interpolated and interpreted as candidate thresholds. The y-axis response was intersected with the regression line and then reflected down to the x-axis stressor, solving for x given the regression equation and y. Results exceeding the observed range of nutrient values and from non-significant regressions ($p>0.05$) were de-emphasized.

Figure 2 illustrates how the nutrient-diatom metric relationship was used in deriving nutrient thresholds. For each site class and each metrics, a meaningful percentile of the reference metric distribution (such as the 25th) was identified. Nutrient values associated with the reference metric values were interpolated by entering the graph from the y-axis, turning at the regression line, and projecting down to the x-axis. This gives one potential threshold value. Other values can be associated with projections from the confidence intervals around the regression line or other reference metric percentiles.

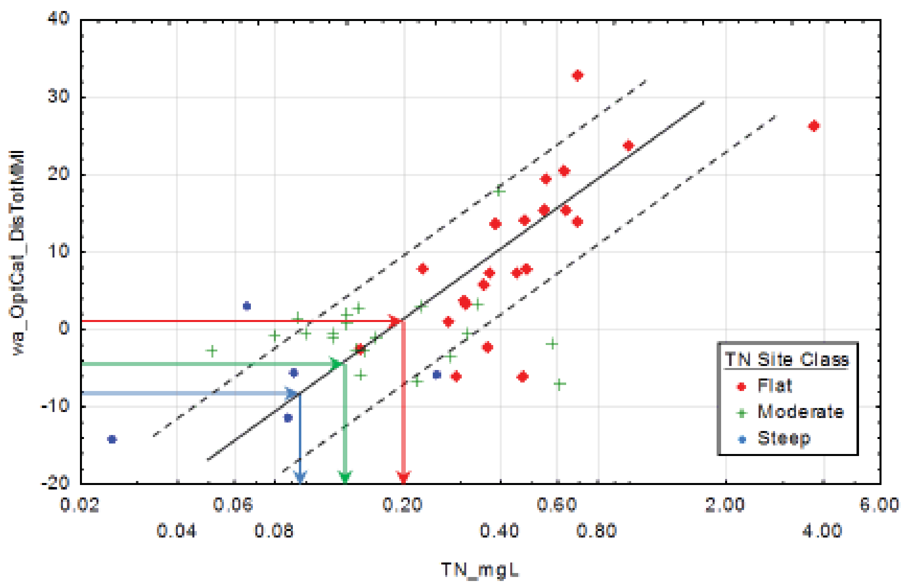


Figure 2. Diatom metric values (weighted average total disturbance multimetric index) against TN concentrations, marked by site class; NRSA data. This figure shows derivation of potential thresholds based on reference percentiles of the metric value and interpolation of nutrient values.

3.5 Change-point Analysis

Change-point analysis with deviance reduction was conducted in R software using the recursive partitioning (rpart) code (Therneau et al. 2013). Output from the change-point analyses included the change-point as well as 95% confidence intervals estimated from a bootstrap re-sampling technique. The plots also included 90% quantile and LOWESS regression lines to allow interpretation of the identified change-points. Results were tabulated and plotted for each site class, nutrient, and response variable. A change-point will be identified whether or not it is a realistic representation of the point at which stressor levels are having critical effects on the response variable. The confidence interval, LOWESS regression, and quantile regression were used to qualify the change-point, which was disregarded if it was considered unrepresentative (Figure 3).

The width of the confidence interval relative to the range of all values indicated precision of the identified change-point. Confidence intervals that included more than half of the stressor scale indicate an imprecise threshold. However, a wide confidence interval alone might not be sufficient reason to disregard a change-point if the LOWESS and quantile regressions suggest an appropriate change-point. The width of the confidence interval was rated as narrow (< ¼ of the entire scale), moderate (¼ to ½ of the scale), or wide (> ½ of the scale).

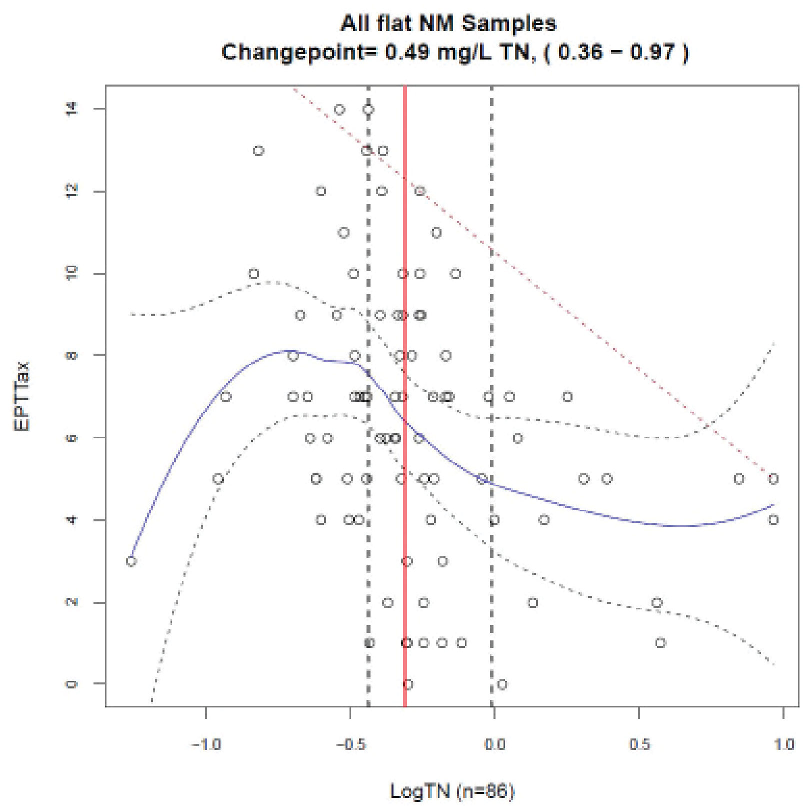


Figure 3. Macroinvertebrate metric values (EPT taxa) against log TN concentrations. The validity of the change-point (verticle solid line) is determined by qualifying the 90% confidence interval (vertical dashed lines), the LOWESS regression line (blue curve) and the 95% quantile regression (dashed sloping line).

The ideal application for change-point analysis is a step-function response. If the LOWESS curve was steep at the change-point, it indicated a step in the response at that point. If the LOWESS fit did not show a visually recognizable change at the identified change-point, then the change-point was de-emphasized. The LOWESS technique (Cleveland 1979) is designed to address nonlinear relationships. LOWESS fits simple models to locally weighted subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point. A bandwidth that considered 75% of the data for smoothing the slope at each data point was used. The LOWESS regression line was used in combination with other indicators of nutrient change-points, primarily as a visual confirmation of changing biological measures at certain nutrient concentrations.

A second check on the general response pattern uses quantile regression to confirm limiting effects of nutrient concentrations on the response variable. Quantile regression is a method for estimating relationships between variables along the upper (or lower) boundary of a distribution of stressor-response data points (Cade et al. 1999). The quantile regression line represents biological potential (plotted on the y-axis) in relation to the stressor of interest (plotted on the x-axis). If limiting factors such as nutrients act as constraints on organisms, then the potential maximum biological condition is observed as a sloping line on a wedge-shaped scatter plot of a biological metric against a nutrient variable. Points that are not along the slope of the wedge represent sites where biological condition is affected by factors not represented on the x-axis. The 90th quantile regression line illustrates the change in the potential biological resource for each increment of disturbance, especially when it is consistent with expectations of trends with increasing nutrients and the LOWESS regression slope at the change-point is in the same direction as the quantile regression line. “R” software (R Core Team 2013) and associated code (quantreg) was used to estimate limited relationships with quantile regression.

3.6 Synthesis of Multiple Thresholds

The frequency distribution and stressor response techniques resulted in multiple candidate thresholds; by stressor, biological assemblage, response measure, site class, and analytical technique. The candidate thresholds were described in the context of strengths and limitations for each analytical method. Results of the multiple methods are expected to agree with each other. When outlier nutrient threshold values were encountered, they were addressed individually and included or excluded from consideration based on analytical integrity (e.g., significance of regressions), feasibility (e.g., within the observed range of values), and corroboration of evidence (e.g., agreement of LOWESS and quantile regression slopes for change-point analysis). The ranges of candidate threshold values and central tendencies were described in narrative, tables, and cumulative distribution function (CDF) curves. The curves illustrated the candidate threshold derived from the reference frequency distribution in relation to candidate thresholds derived through stressor response techniques (Figure 4).

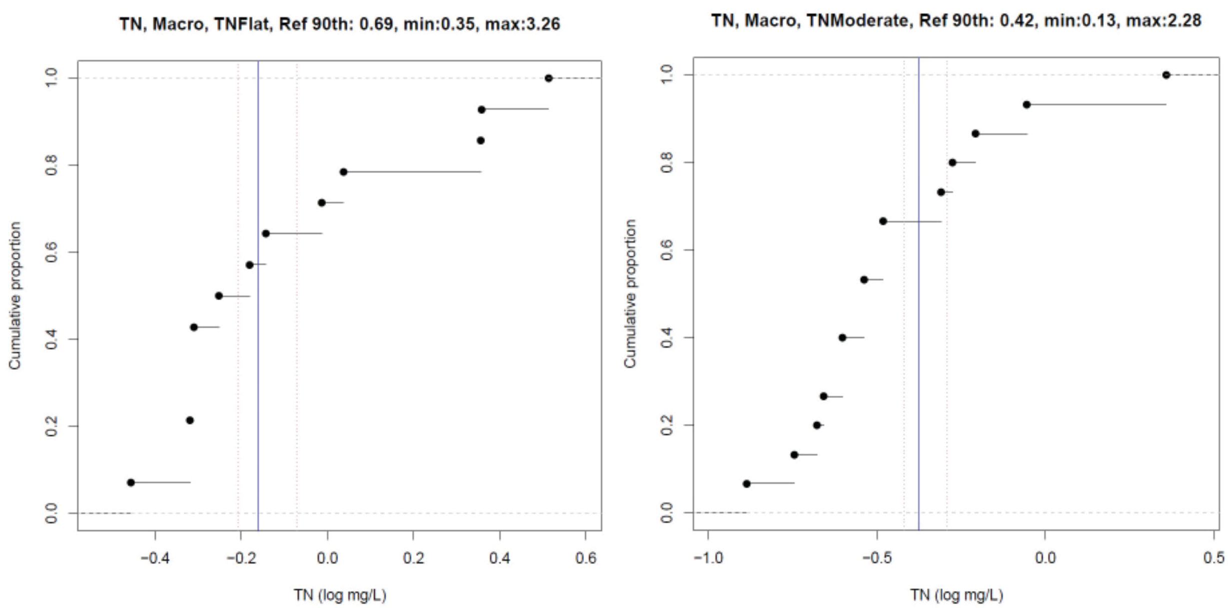


Figure 4. Examples of a CDF curves showing candidate TN threshold values derived from reference frequency distribution (vertical solid line with confidence limits) and from multiple stressor-response analyses of macroinvertebrate metrics using change-point and regression interpolation (points along the curve). The log values in the x-axis are back transformed to mg/L in the titles. These graphs show results for two site classes: TN Flat (left) and TN Moderate (right).

In the example, each point represents a specific threshold value related to particular analytical technique or metric in the tabulated results. In the TN Flat site class, the reference-derived value falls near the 0.58 cumulative value (left graph in Figure 3). If the threshold is established at the reference-derived value, it would higher than ~58% of the stressor-response candidate thresholds. About 70% of the stressor-response-derived values are at or below the reference-derived value in the moderate site class (right graph). On the logarithmic scale, the stressor-response thresholds in the upper right portion of the graphs are much higher than the others, especially those that are to the right of the points with long horizontal leaders.

4.0 Results

4.1 Reference Sites and Classification

The reference site analysis and disturbance gradient designations resulted in 20% of sites identified as least disturbed reference sites (Table 10, Appendix F). Another 11% were designated as near-reference. This is a reasonable proportion of reference sites because sites with potential for least-disturbed reference status were targeted when selecting sites for sampling and analysis. Smaller percentages of sites were designated as stressed (7%) or extremely stressed (6%). The remaining sites were designated as “other”, having moderate levels of disturbance.

Most of the designations (83%) assigned by numeric site criteria based on GIS analysis of land use coverage and human activity in the catchments were confirmed during the NMED review. Of the 114 designations that were not confirmed by NMED, 70 were reductions from a better category to a worse one. When NMED suggested changes that were only one category (e.g., NearRef -> Ref), further review was not required. The 29 sites that changed by more than one category (e.g., Other -> Ref) were reviewed by the workgroup using site history, known stressors, and aerial photos to resolve designations for each site. Site reviews used landscape and historical input, not water quality information. However, there were three sites that were downgraded from ‘NearRef’ to ‘Other’ because high NO_3NO_2 in relation to TKN indicated nutrient additions.

Reference sites were sought in all regions of New Mexico using the three site classes established for stream sediment analysis (Jessup et al. 2010) to compare the distribution of reference sites throughout New Mexico. As expected, mountainous regions had a greater percentage of reference sites in comparison to either the foothills or xeric areas. Conversely, stressed sites were more common in the xeric areas compared to the foothills or mountains.

Table 10. Reference site designations by reference status and sediment region.

Region	N	Ref	NearRef	Strs	XStrs	%Ref	%Strs
Mountains	225	66	29	10	7	29.3	7.6
Foothills	197	33	19	12	4	16.8	8.1
Xeric areas	240	14	23	25	30	5.8	22.9
Total	662	113	71	47	41	20.5	13.3

For site classification, the data frame contained 542 wadeable stream sites with complete landscape analysis and nutrient concentrations, summarized per site. Of the 662 sites identified for GIS analysis, 120 were in non-wadeable rivers, were intermittent, were duplicates of other analyzed sites, or lacked critical data. Preliminary explorations (Appendix G) indicated that we could pool nutrient data across data sources (NMED, NRSA, and WSA) and that the reference and near-reference sites should be combined for the remaining classification exercises. Existing classification schemes based on level 2 ecoregions (Stoddard et al. 2005) or sediment regions (Jessup et al. 2010) showed insufficient separation of nutrient distributions and a new analysis was warranted.

Phosphorus

Phosphorus values were first partitioned by longitude in the CART models both with and without categorical variables (Figure 5). Cross validation for CART is calculated with random permutations, leading to different results for each run. In most runs, cross validation indicated overfitting after the first split based on longitude. In other iterations, several splits could be included without overfitting and these are shown in the figure. The additional splits were based on average land slope, latitude, and precipitation. To interpret the correct number of splits, TP values were plotted with the variables used in the CART analysis. Excessive partitioning of the data would be conceptually (if not statistically) overfitting the data and many or small site classes would add unnecessary complexity. The random forest analysis suggested that the most important classifying variables were longitude (importance measure = 3.04), land slope (2.73), latitude (2.32), and precipitation (2.09). These relative importance measures using continuous variables did not change when repeating the analysis with categorical variables included.

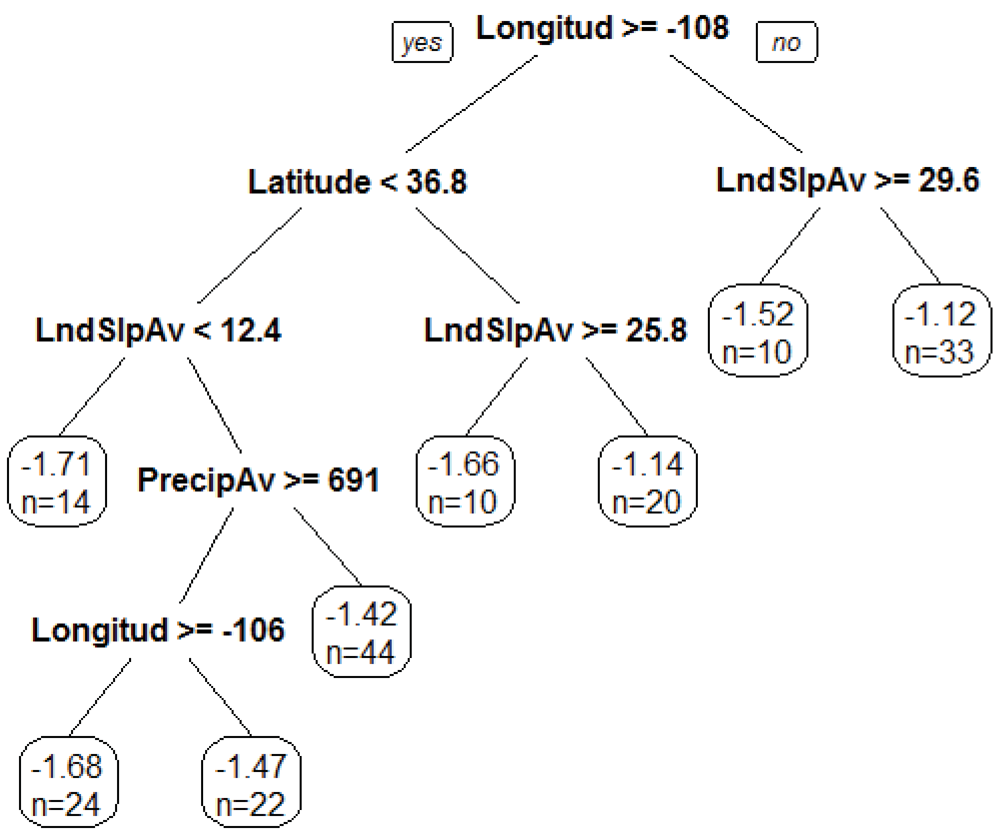


Figure 5. Classification and Regression Tree (CART) for average total phosphorus (TP). At the first split, 134 of 177 reference and near-reference sites east of longitude -108.1 were partitioned to the left of the tree. Additional splits in the data were based on latitude, land slope and precipitation. At the end of each branch the average TP concentration (log mg/L) and number of sites are displayed.

Longitude was consistently an important variable, occurring in the first CART split and having the greatest importance in the random forest analysis. The split was defined at longitude -108.13, which includes western ecoregions (Colorado Plateaus [ecoregion 20], Arizona/New Mexico Mountains [23], Madrean Archipelago [79], and the Chihuahuan Desert [24]). The western parts of these ecoregions had higher TP than the eastern parts (Figure 6). Reference sites in the Arizona/New Mexico Plateau (22) were either east of longitude -108 or mostly drained other ecoregions. In the Arizona/New Mexico Mountains, longitude -108 is an approximate watershed boundary between the Rio Grande and Gila River basins.

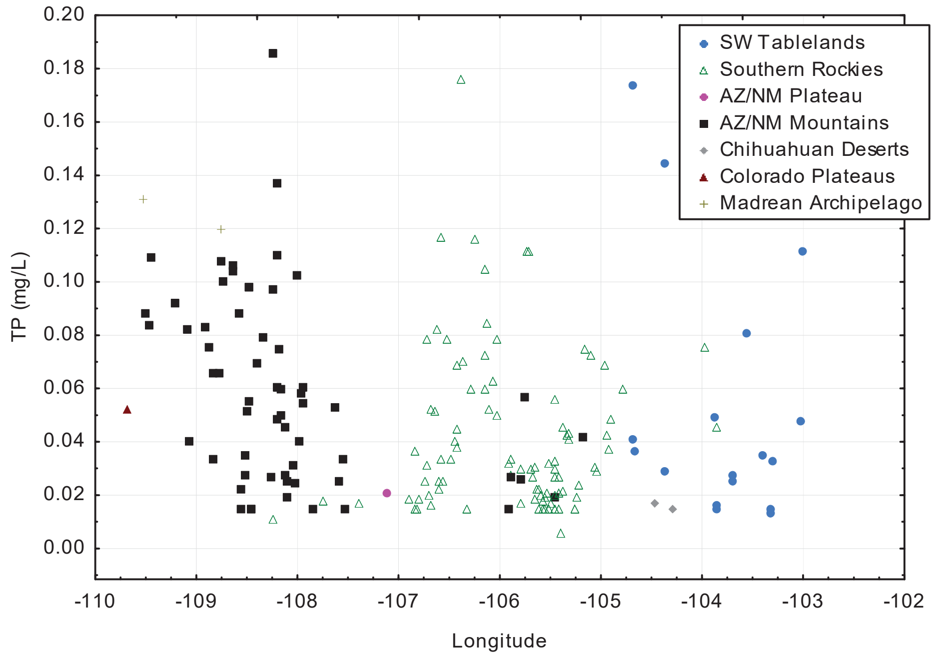


Figure 6. Total Phosphorus (TP) average values per reference or near-reference site in relation to longitude, showing the dominant ecoregion of the site catchment.

CART analysis split the eastern data by latitude, but with a latitude threshold that was so far north in the state (latitude 36.8 passes north of Ensenada, NM) that the threshold did not make sense for site classification in New Mexico. Land slope was the first split of CART analyses conducted separately for sites in western or eastern classes. Steeper sites have lower TP, in general (Figure 7). A CART analysis forcing land slope as the classification variable in all sites resulted in a split threshold of 29%.

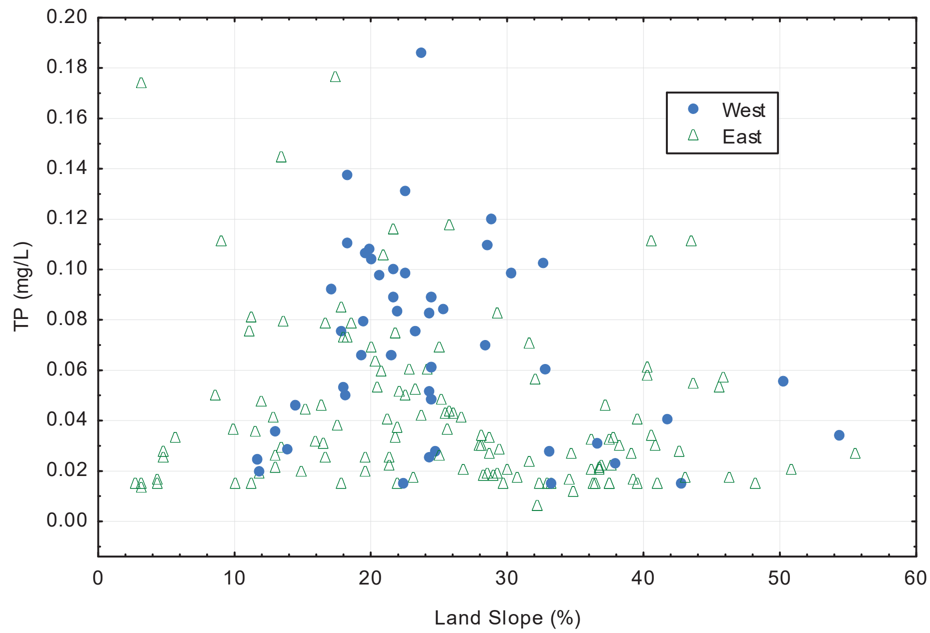


Figure 7. Total Phosphorus (TP) average values per reference or near-reference site in relation to average land slope in the catchments, showing East and West regions (longitude -108).

NMED reviewed the initial classification analysis resulting in classes defined by longitude and land slope. The longitude split appeared to be driven by the large number of reference sites in the higher background TP watersheds in SW New Mexico. Higher background TP was suspected of being related to volcanic geology, but the geologic designations alone could not explain differences in reference TP. While all of the high TP reference sites were in volcanic regions, other volcanic formations did not have high background TP. Instead, specific basins (8-digit HUCs) were identified with high TP in reference sites. These watersheds correspond to those shown to have high soil TP (Woodruff et al. 2015).

The TP High-Volcanic site class was defined for regions with relatively high background TP, most of which were associated with volcanic geology (Table 11). Flat western sites of the initial classification had more homogenous reference TP values when non-volcanic basins were removed from this group which became the TP High-Volcanic group with the addition of other volcanic and high soil TP basins. The TP High-Volcanic site class includes the following basins: the Upper Gila, the Upper Gila-Mangas, the San Francisco, the Mimbres, and the San Antonio/Conejos. The San Antonio/Conejos is the only basin that is not in southwest NM. It is a volcanic region along the central section of the northern border of New Mexico. The following smaller basins (12-digit HUCs) were excluded though they are in the Upper Gila basin: Diamond, Taylor and Beaver Creeks (HUCs 150400010404, 150400010406, 150400010402, 150400010403, 150400010305, and 150400010302). The Jemez basin was suspected of being part of this class, but was not because background TP levels were not as high as in other TP High-Volcanic sites.

Sites not in the TP High-Volcanic class were separated into two classes based on 29% average catchment land slope. The TP Steep class has sites with slopes greater than 29% and background TP concentrations that were the lowest of the three classes. The TP Flat-Moderate class has flatter landscapes, though three basins with marginally flat sites (<31.8% land slope) were included because background TP concentrations were higher than typical TP Steep sites. These exceptions included drainages in the Vallecitos, Pajarito and Sulfer/Redondo basins (HUCs 130202020204, 130202010204, and 130202020202).

These three classes had significantly different TP values (Figure 8) based on the non-parametric Kruskal-Wallis test ($p<0.01$ for all comparisons).

Table 11. Site classes for TP and TN.

TP High-Volcanic –The class includes all sites in the San Antonio and Conejos, the Upper Gila, Upper Gila-Mangas, San Francisco, and Mimbres basins. In the Upper Gila basin, it excludes sites in the Diamond, Taylor and Beaver Creek sub-basins (HUCs 150400010404, 150400010406, 150400010402, 150400010403, 150400010305, and 150400010302).
TP Flat-Moderate - This class includes all sites less than or equal to 29% average land slope and not in the TP High-Volcanic site class. It also includes sites in three drainages of the Jemez basin, the Vallecitos, Pajarito, and Sulpher/Redondo sub-basins (HUCS 130202020204, 130202010204, and 130202020202).
TP Steep - The Steep class includes all sites with average land slopes greater than 29% and not in the TP High-Volcanic site class.
TN Flat - TN Flat sites have average catchment land slopes less than 15%
TN Moderate - TN Moderate sites have average catchment land slopes from 15% to 32%
TN Steep - TN Steep sites have average catchment land slopes greater than 32%

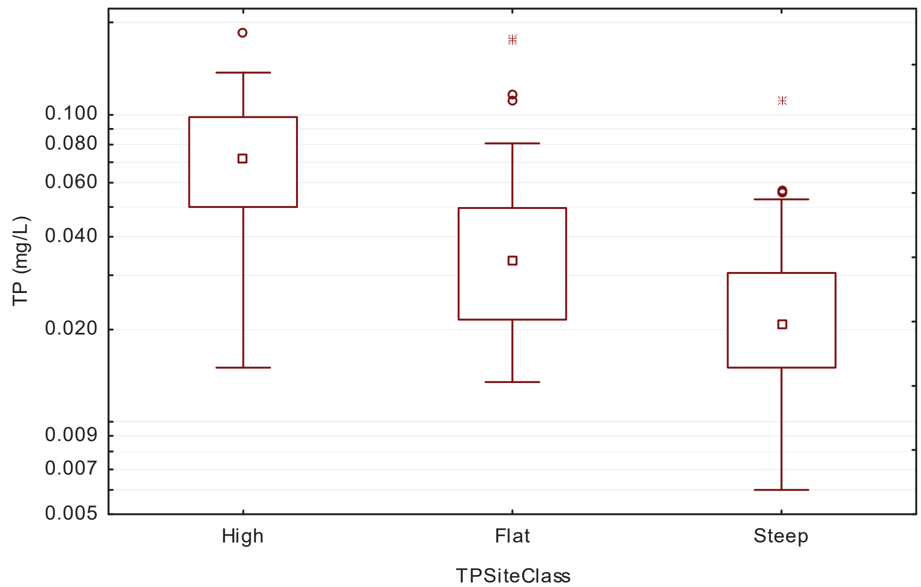


Figure 8. Total Phosphorus (TP) concentrations in reference or near-reference sites by potential site classes for TP. Sample sizes are 55, 76, and 48, in the order displayed.

Nitrogen

Similar to phosphorus, total nitrogen values were first partitioned by longitude in the CART models both with and without categorical variables (Figure 9). Importance coefficients in the random forest analysis were as follows: longitude (1.92), land slope (1.76), precipitation (1.57), latitude (1.44), and temperature (1.28). Before the tree was pruned to avoid over-fitting, land slope, longitude, and precipitation appeared in additional branches to further classify the 131 western warmer sites.

The split for longitude was at -105.2, with higher TN values in the east (Figure 10). Eastern ecoregions include the Southwestern Tablelands, High Plains, eastern portion of the Chihuahuan Desert, and small parts of Southern Rockies and Arizona/New Mexico Mountains. Of the 32 eastern reference sites, 25 were in the Southwestern Tablelands, 5 were in the Southern Rockies, and 2 were in the Arizona/New Mexico Mountains. The 5 Southern Rockies sites had TN values that were similar to other Southern Rockies sites that were just west of the longitudinal threshold. Because the ecoregions align with the longitudinal threshold and make more ecological sense in explaining nutrient conditions than a line of longitude, the 'Eastern' sites were defined by ecoregion, including Southwestern Tablelands and the eastern portion of the Chihuahuan Desert (east of longitude -105). The High Plains would be included in the Eastern class if samples were collected there, but they weren't.

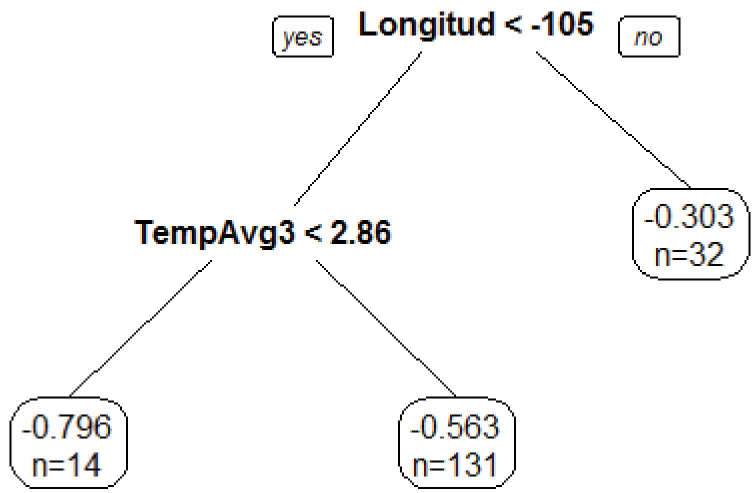


Figure 9. Classification and Regression Tree (CART) for average Total Nitrogen (TN). At the first split, 145 of 177 reference and near-reference sites west of longitude -105.2 were partitioned to the left of the tree. An additional split was based on air temperature. At the end of each branch the average TN concentration (log mg/L) and number of sites are displayed.

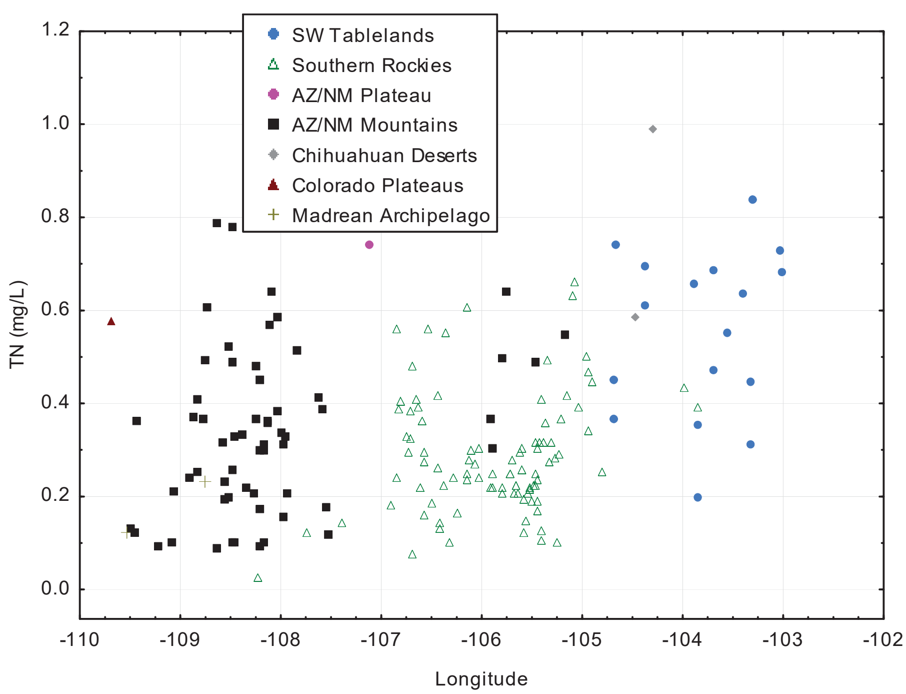


Figure 10. Total Nitrogen (TN) average values per reference or near-reference site in relation to longitude, showing the dominant ecoregion of the site catchment.

Cross validation indicated overfitting after the second split in the CART analysis, which included average air temperature splitting the western group. The air temperature split resulted in a small (N = 14) low temperature group with low TN. The random forest analysis showed that importance was lower for temperature compared to other variables and this small group was not valid as a site class.

Land slope was explored as a classification variable because it was important in the random forest analysis. A CART analysis with all sites forcing only land slope in the model resulted in 2 splits at 15% and 32%. The flattest landscapes were mostly associated with eastern sites, though some western sites had flat slopes and similar TN values to the eastern sites with flat slopes (Figure 11). The eastern streams in the moderate slope category had TN values in the same range as western streams in that slope category. In the steep category, there were three notable outliers with higher TN values that could not be explained by the classification variables (77Diamon033.2, 57RBonit061.1, and 77Turkey001.8).

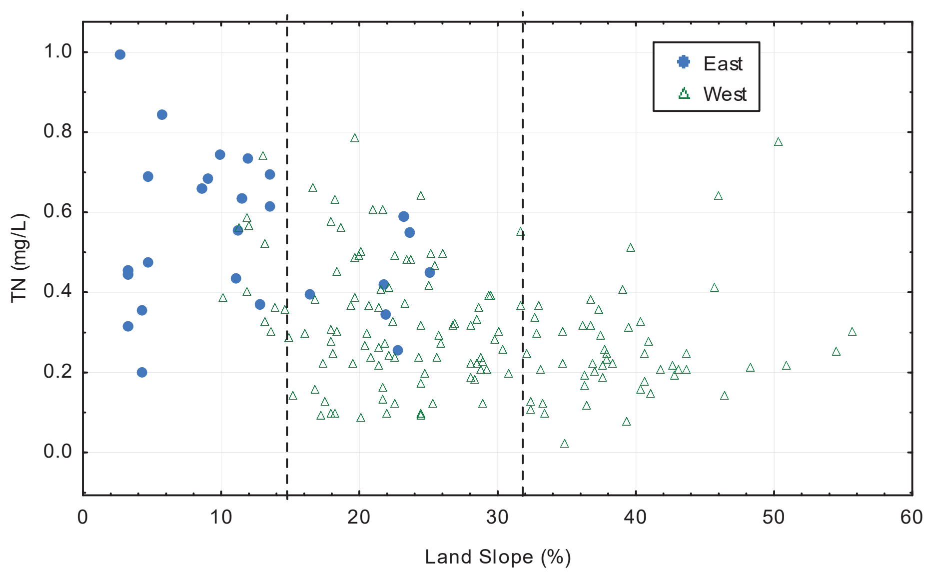


Figure 11. Total Nitrogen (TN) in relation to land slope, showing east and west designations derived from the first CART split (longitude -105).

The western streams with the flattest landscapes were represented by only 12 sites. Because this is a small group and because land slope appears to partition TN values as well as or better than longitude and land slope, classes were based on land slope alone. Thresholds of 15% and 32% defined TN Flat, TN Moderate, and TN Steep classes, regardless of longitude (Table 11). This classification scheme resulted in distinct TN values within the classes (Figure 12). The TN values were significantly different based on the non-parametric Kruskal-Wallis test. The differences in relation to the Flat class ($p<0.001$) were greater than the difference between the Moderate and Steep groups ($p=0.03$).

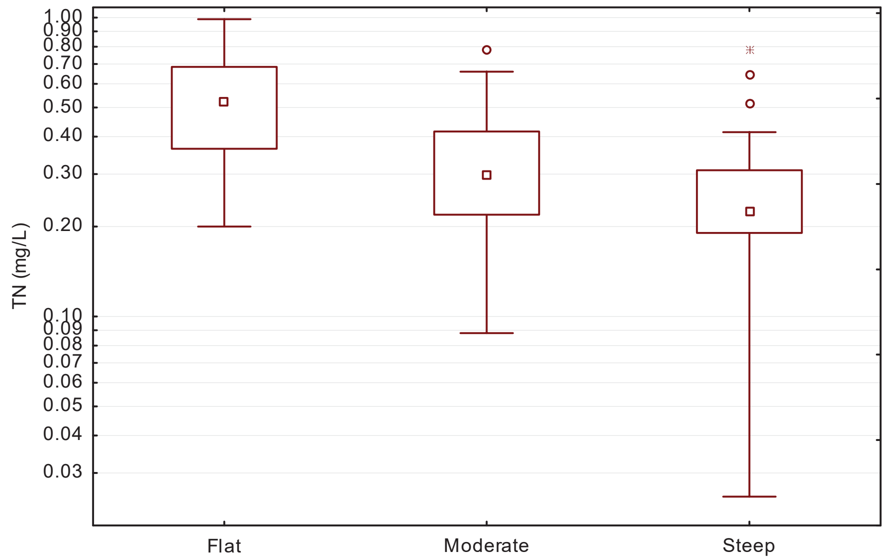


Figure 12. Total Nitrogen (TN) concentrations in reference and near-reference sites by potential site classes. Sample size for TN Flat, TN Moderate, and TN Steep site classes are 31, 95, and 51, respectively.

Considerations for application

NO_3NO_2 was suspected to be dominant in the sites with high TN concentrations. However, this was not necessarily true. In the steep TN class, both NO_3NO_2 and TKN had reference and near-reference concentrations generally below 0.20 mg/L. In the TN Flat class, TKN was higher (upper quartile = 0.56 mg/L in reference and near-reference sites) while NO_3NO_2 was still low (generally below 0.20 mg/L).

Different site classes for TN and TP were not anticipated when classifying sites to partition nutrient variability. In the independent analyses for each nutrient, similar classification variables, longitude and land slope, were identified. Though identical site classes for TN and TP were attempted (Appendix G), the nutrient specific classes were more precise and are appropriate for application of nutrient thresholds.

Maps of the site classes show that land slopes are variable within the ecoregions (Figures 13 and 14). For TP (Figure 13), most of the TP Steep sites are in the eastern arm of the Southern Rockies (Sangre de Cristo Mountains). TP Steep sites are also found in the Black Range and Sacramento Mountains of the Arizona/New Mexico Mountains ecoregion. The TP High-Volcanic sites are mostly in the San Francisco Mountains northwest of Silver City. The TP Flat-Moderate sites are throughout the state. For TN (Figure 14), the TN Steep sites are in the same general area as the TP Steep sites, except that the TN Steep sites extend further west in the San Francisco Mountains where they are recognized as the TP High-Volcanic class for TP. The TN Flat sites (<15% land slopes) are mostly in the xeric ecoregions.

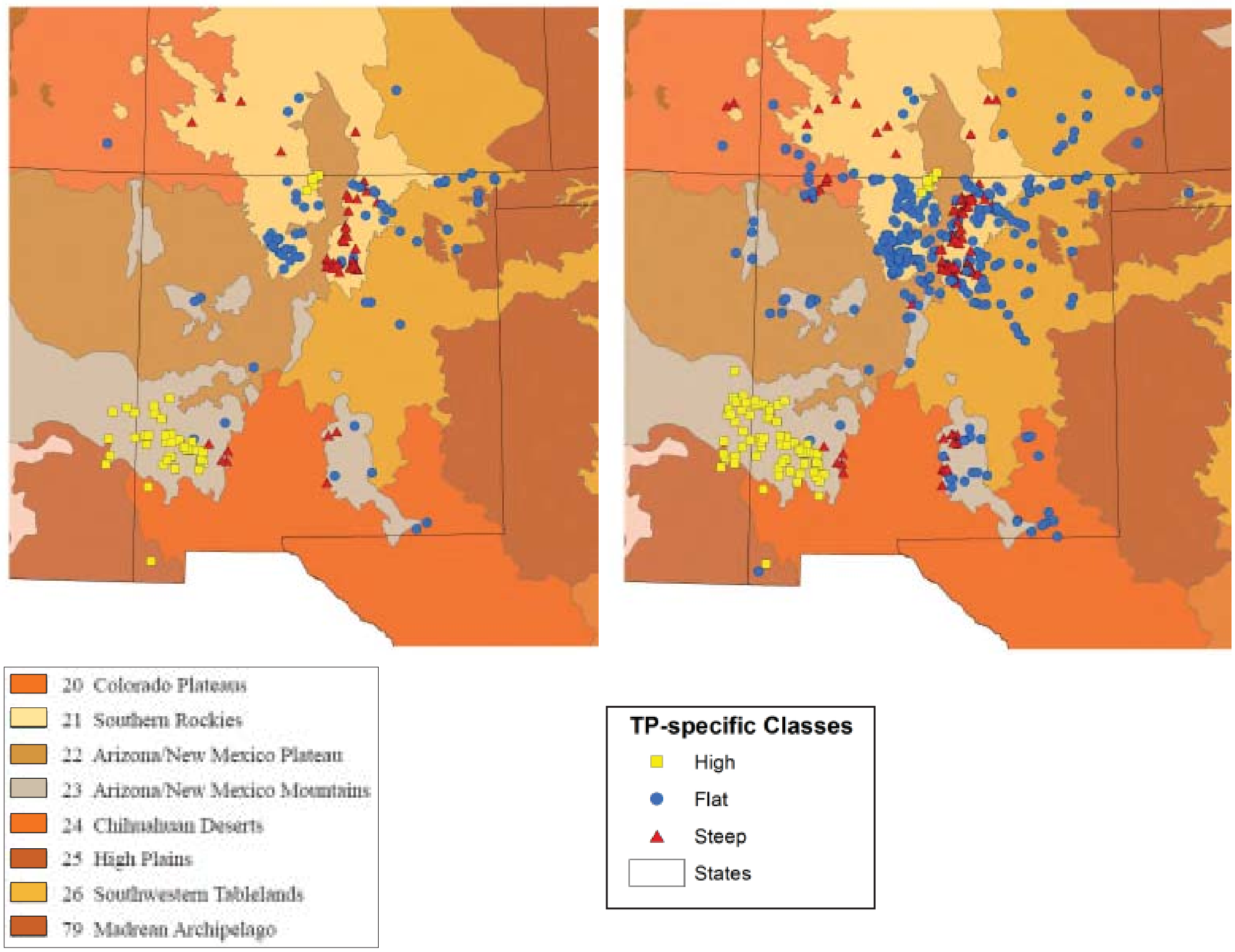


Figure 13. . Sites in the ecoregions of New Mexico showing reference and near-reference sites (left) and all sites (right), marked by the TP-specific site classes.

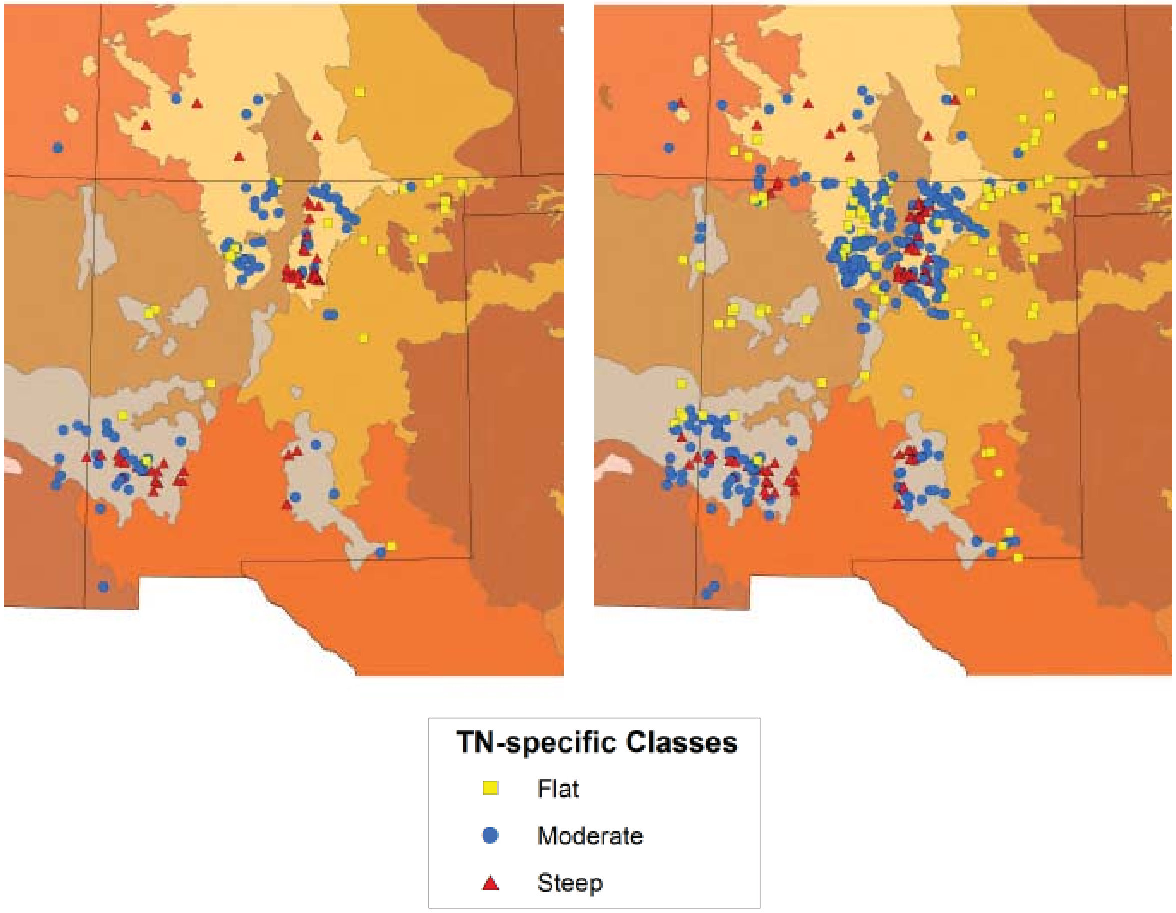


Figure 14. Sites in the ecoregions of New Mexico showing reference and near-reference sites (left) and all sites (right), marked by the TN-specific site classes. Ecoregions are as in Figure 13.

4.2 Frequency Distributions

The reference condition analysis resulted in quantiles from frequency distributions that were considered as candidate thresholds for TN and TP in data subsets by site class. Within sites, the long term medians of TN and TP concentrations were used instead of an average to reduce the influence of temporary extreme values. Benthic chl-a concentrations were evaluated in the TP classes and had only one observation per site. Among sites and within site classes, the median, 75th, 80th, 85th, and 90th quantiles of the concentrations were presented to characterize the reference and near-reference sites (combined) (Table 12). NMED chose to emphasize the 90th quantile to represent candidate thresholds for nutrients. Quantile selection for candidate threshold determination is dependent upon the data, and the certainty one has that they accurately reflect reference conditions. The quantile selected must be protective of the designated uses. The U.S. EPA generally recommends using the 75th quantile of reference sites (USEPA 2000). For this analysis, there was a high level of certainty in reference site selection.

The 75th quantile of median site values was considered and rejected because this value did not include naturally enriched systems in the reference data set. In addition, the 90th quantile was more closely aligned with the benthic macroinvertebrate and diatom change point analyses, and is hence assumed protective of the applicable designated aquatic life use(s). Therefore, the 90^{th} quantile was preferred to represent candidate thresholds in New Mexico streams.

Table 12. Frequency distribution statistics for median TP, TN, and benthic chl-a concentrations in valid reference and near-reference sites. The preferred candidate threshold (90^{th} quantile) is shown in bold-type.

	TP (mg/L)			TN (mg/L)			Chl-a (ug/cm^2)		
Quantile	Lower 90% CI	Value	Upper 90% CI	Lower 90% CI	Value	Upper 90% CI	Lower 90% CI	Value	Upper 90% CI
	TP High-Volcanic (N=55)			TN Flat (N=30)			TP High-Volcanic (N=25)		
50th	0.049	0.058	0.071	0.38	0.47	0.56	1.05	2.41	3.52
75th	0.072	0.084	0.09	0.55	0.61	0.67	2.74	4.41	6.01
80th	0.08	0.088	0.104	0.56	0.62	0.7	3.21	5.00	6.65
85th	0.084	0.092	0.106	0.59	0.65	0.84	3.7	5.66	8.5
90th	0.089	**0.105**	0.114	0.62	**0.69**	0.85	4.38	6.39	11.9
	TP Flat-Moderate (N=76)			TN Moderate (N=96)			TP Flat-Moderate (N=42)		
50th	0.016	0.025	0.033	0.23	0.25	0.28	1.64	2.24	3.18
75th	0.034	0.041	0.05	0.33	0.35	0.37	3.24	4.93	14.95
80th	0.036	0.048	0.058	0.35	0.37	0.41	3.98	8.06	19.03
85th	0.043	0.054	0.061	0.36	0.40	0.45	4.09	15.68	24.15
90th	0.051	**0.061**	0.069	0.38	**0.42**	0.51	8.38	20.98	25.67
	TP Steep (N=48)			TN Steep (N=53)			TP Steep (N=14)		
50th	0.015	0.015	0.015	0.18	0.20	0.21	0.73	1.66	3.31
75th	0.015	0.015	0.018	0.21	0.23	0.27	1.74	3.33	13.53
80th	0.015	0.016	0.023	0.22	0.25	0.3	1.89	7.29	17.47
85th	0.015	0.018	0.035	0.24	0.28	0.33	2.42	13.22	23.86
90th	0.016	**0.030**	0.053	0.26	**0.30**	0.34	2.91	13.51	23.86

To illustrate the validity of using reference quantiles to derive thresholds for the New Mexico data sets, distributions of TP, TN and benthic chl-a were plotted by site class and reference status (Figures 15-23). For TN and TP, reference distributions were similar to near-reference distributions within each site class. Nutrient concentrations generally increased with increasing disturbance. The stressed site distributions were similar to the reference distributions only for TP in the TP High-Volcanic site class (Figure 15). Stressed and Highly Stressed categories included only 3 and 2 sites, respectively. In the other classes, the stressed and extremely stressed sites had at least 50% of the values greater than the 75^{th} quantile of the reference and near-reference site values. For benthic chl-a (Figures 21-23), median values in Stressed and Highly Stressed categories were consistently higher than medians in Reference and Near-Reference categories, though the stressed categories were represented by fewer than five samples in all but the TP Flat-Moderate class.

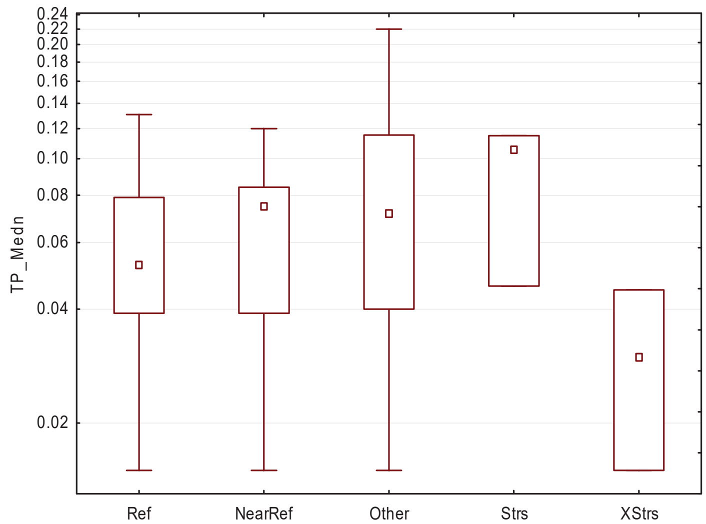


Figure 17. Site median TP value distributions along the disturbance gradient for sites in the TP High-Volcanic site class.

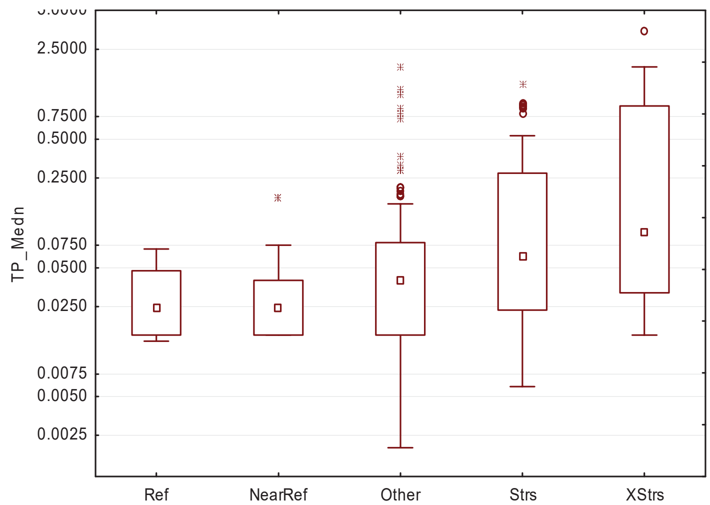


Figure 16. Site median TP value distributions along the disturbance gradient for sites in the TP Flat-Moderate site class.

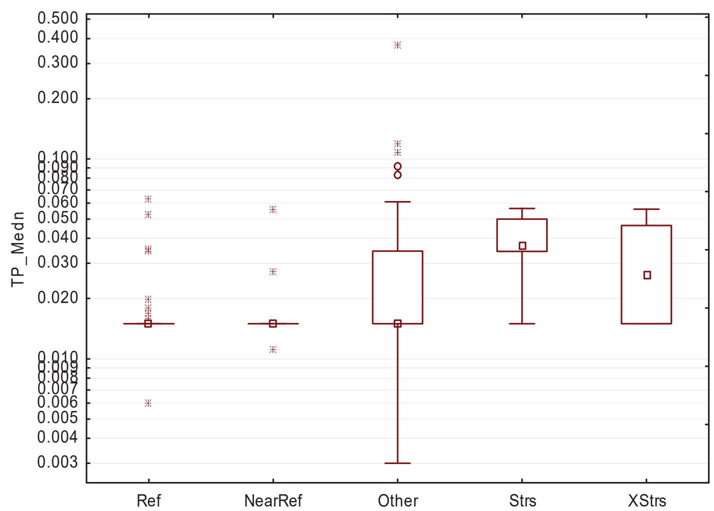


Figure 15. Site median TP value distributions along the disturbance gradient for sites in the TP Steep site class.

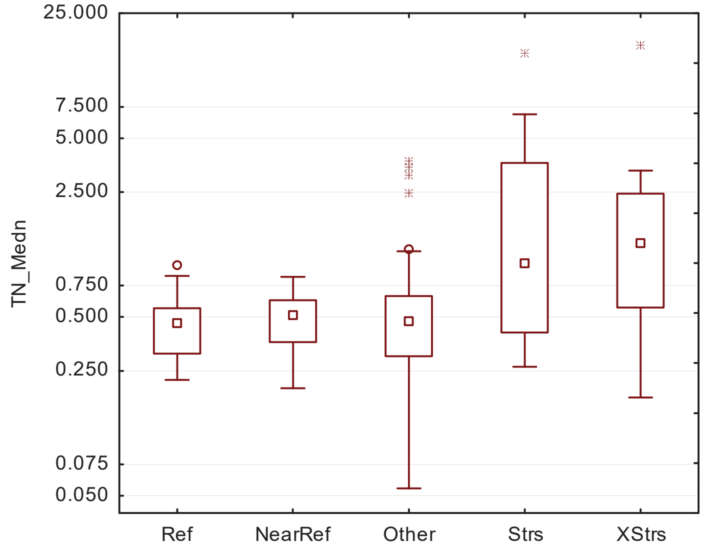


Figure 18. Site median TN value distributions along the disturbance gradient for sites in the Flat site class.

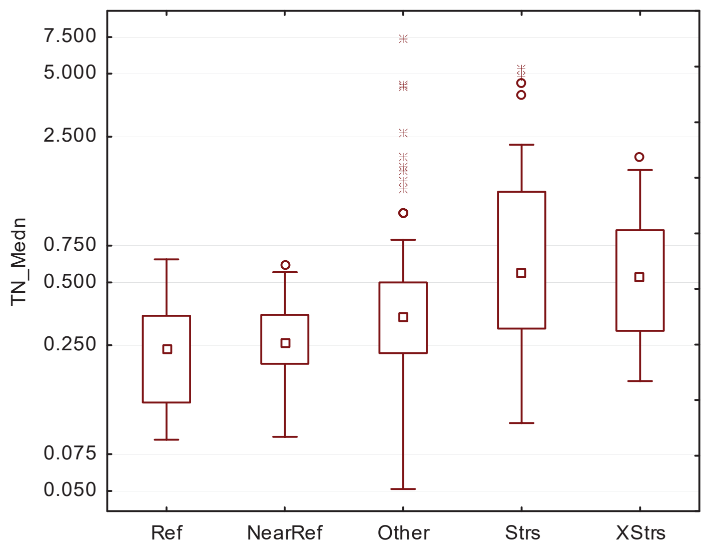


Figure 19. Site median TN value distributions along the disturbance gradient for sites in the Moderate site class.

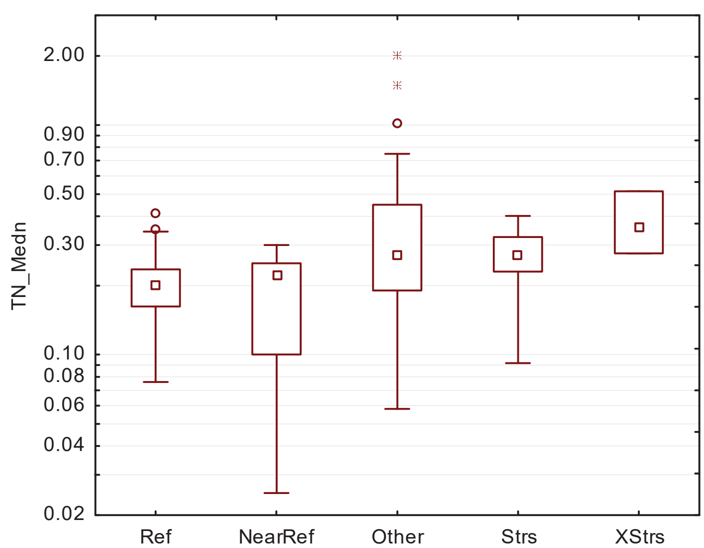


Figure 20. Site median TN value distributions along the disturbance gradient for sites in the Steep site class.

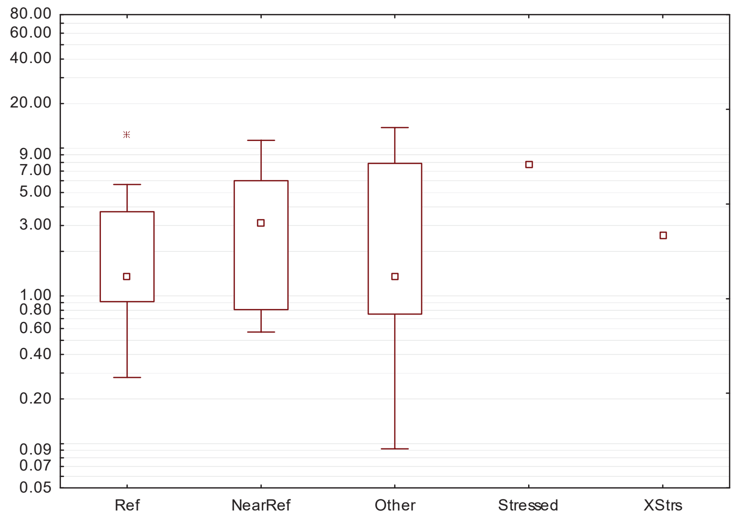


Figure 21. Chl-a distributions along the disturbance gradient for sites in the TP High Volcanic site class.

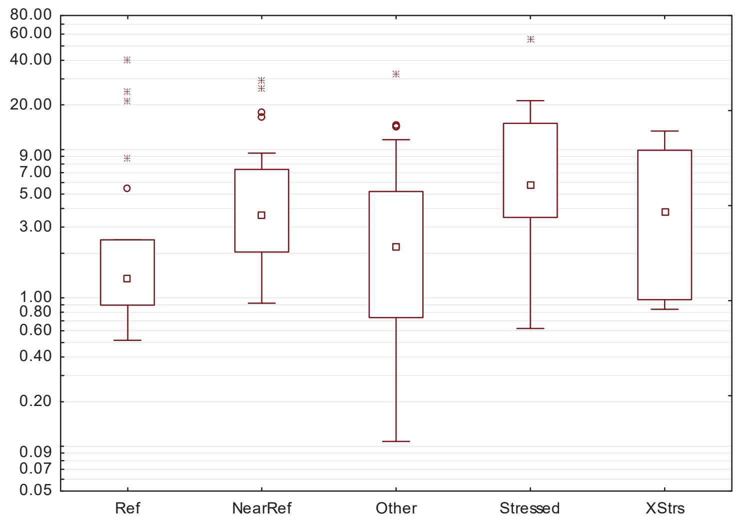


Figure 22. Chl-a distributions along the disturbance gradient for sites in the TP Flat-Moderate site class.

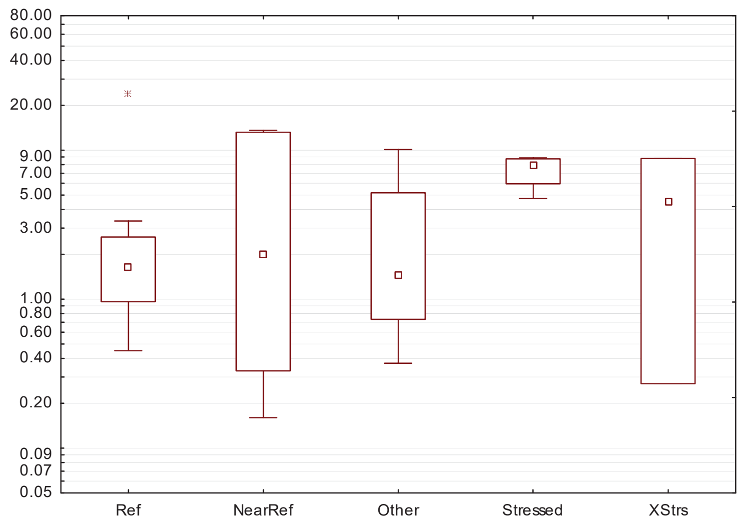


Figure 23. Chl-a distributions along the disturbance gradient for sites in the TP Steep site class.

Quantiles of DeltaDO and Pmax4hr diel DO measures in reference and near-reference sites were similar in the High-Volcanic and TP Flat site classes and were relatively lower in the TP Steep site class (Table 13). DeltaDO and Pmax4hr increased with increasing stress in the streams, especially in the TP Flat site class (Figure 24).

Table 13. Frequency distribution statistics for diel DO statistics in valid reference and near reference sites. The preferred candidate threshold (90th quantile) is shown in bold-type.

Delta DO	High-Volcanic			Flat			Steep		
Statistic	50th	75th	90th	50th	75th	90th	50th	75th	90th
lower 90% CI	1.93	2.17	3.13	1.22	2.28	3.52	1.10	1.10	1.40
quantile	2.17	3.27	**5.02**	1.82	3.06	**4.08**	1.13	1.57	**1.79**
upper 90% CI	3.03	4.29	7.24	2.50	3.98	7.26	1.57	2.37	2.37
Prod4hr	High-Volcanic			Flat			Steep		
Statistic	50th	75th	90th	50th	75th	90th	50th	75th	90th
lower 90% CI	0.176	0.331	0.460	0.148	0.296	0.493	0.095	0.105	0.126
quantile	0.304	0.501	**0.635**	0.208	0.501	**0.682**	0.105	0.186	**0.284**
upper 90% CI	0.439	0.648	0.720	0.393	0.678	1.200	0.196	0.490	0.490

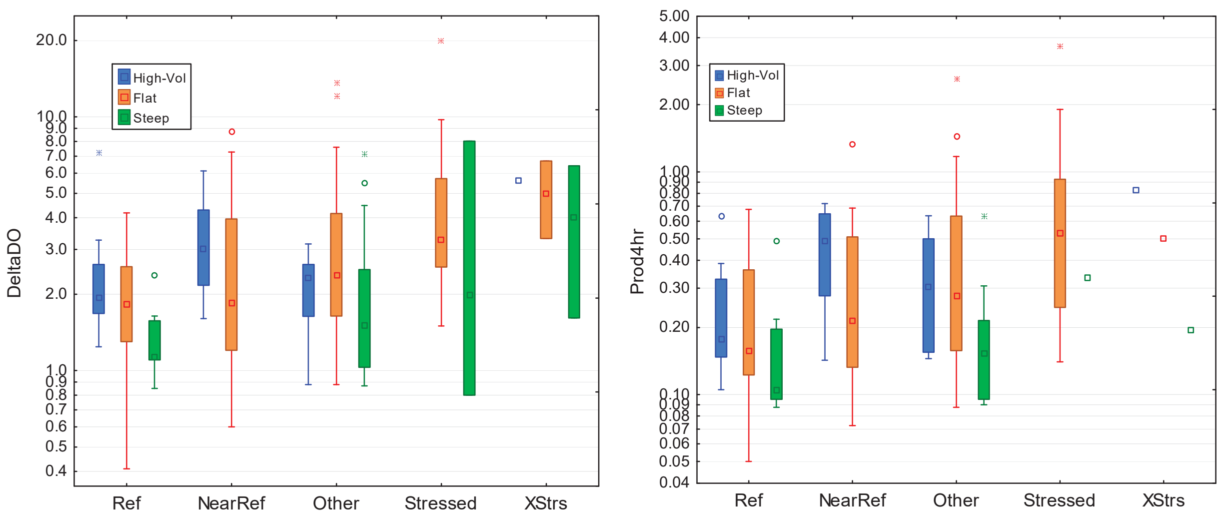


Figure 24. Distributions of Delta DO and maximum 4-hour productivity in disturbance categories and TP site classes.

4.3 Correlations and Interactions

The site classes were a good starting point for developing nutrient expectations though there was variability remaining within the site classes. To strengthen understanding of the primary linkages in the conceptual model (nutrients - chl-a, chl-a - DO dynamics, nutrients - diatom metrics, and DO - macroinvertebrate metrics), each relationship was explored in further detail to determine whether consistent modifiers could be factored out to refine the general relationship. The explorations included correlation analysis, multiple regression, CART, random forest, and graphic displays. Nutrient concentrations collected within 30 days of the response variable were averaged. Censored data were assigned a value of half the detection limit and retained for these analyses.

Distributions statistics for TN, TP and frequently collected water quality variables (grab samples) characterize the perennial wadeable streams in New Mexico (Table 14). The water quality variables include NO3NO2, TKN, pH, conductivity (EC), turbidity, temperature, and DO. Spearman rank correlation analysis for TN and TP and water quality variables shows positive relationships between the nutrients and between nutrients and other water quality measures (Table 15). Positive correlations with pH and EC were stronger for TN than for TP. DO and pH were positively correlated with each other over all sites ($r = 0.22$, $p<0.05$) and in all site classes. Turbidity and temperature had stronger correlations with TP than with TN. The correlation between DO at the time of sampling was negative with TP and positive with TN. Diel DO measures were examined in a subset of sites (see Section 4.3.2). Diel DO measures are more reliable for correlation analysis because they not subject to the high daily variations that are inherent to instantaneous grab data.

Table 14. Distribution statistics for nutrients and water quality variables in stream sites.

	Valid N	Minimum	Lower Quartile	Mean	Median	Upper Quartile	Maximum
TN (mg/L)	538	0.03	0.22	0.59	0.33	0.54	16.44
TP (mg/L)	542	0.002	0.015	0.099	0.036	0.072	3.420
NO3NO2 (mg/L)	441	0.05	0.05	0.31	0.07	0.16	14.00
TKN (mg/L)	444	0.05	0.17	0.41	0.27	0.49	6.55
pH (su)	476	5.6	7.8	8.0	8.1	8.3	10.7
EC (uS/cm)	476	32	159	653	318	656	9195
Temperature (°C)	477	3.9	11.8	15.0	14.7	17.8	63.5
DO (mg/L)	439	3.6	8.0	8.6	8.7	9.2	17.7
Turbidity (NTU)	472	0.0	1.1	20.2	4.2	13.2	1184

Table 15. Spearman rank correlation coefficients for nutrients and water quality variables in all sites and in site classes. Significant correlations (p<0.05) are marked with an asterisk (*).

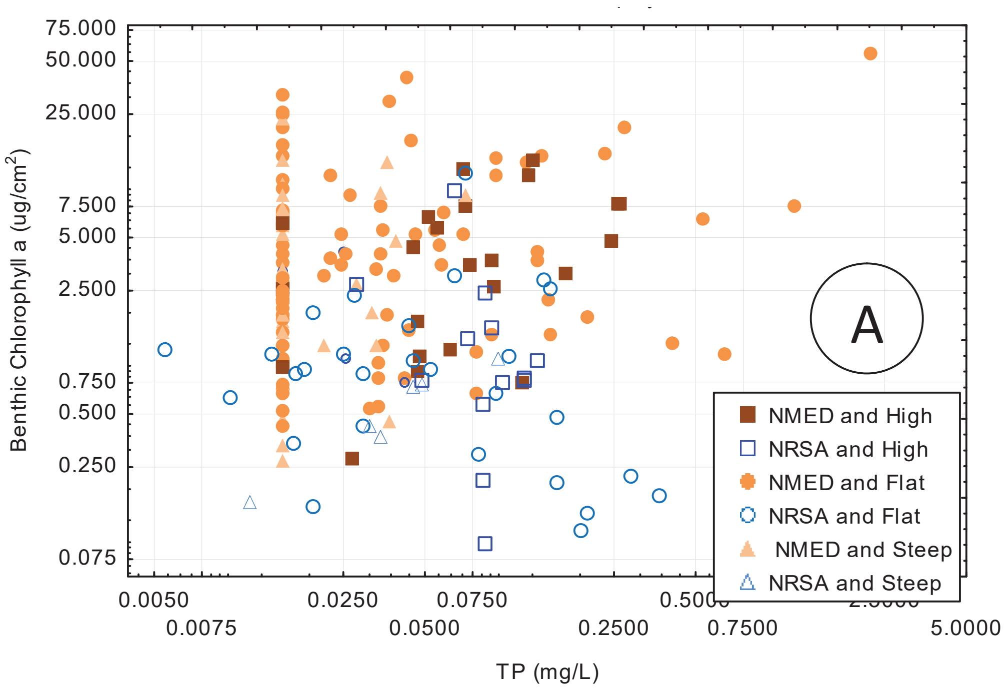
Nutrient	Data SubSet	TN	TP	NO3NO2	TKN	pH	EC	Turb	Temp	DO
TP	All Classes	0.39*		0.13*	0.54*	0.12*	0.13*	0.48*	0.24*	-0.23*
	TP High Volc.	0.38*		0.35*	0.51*	0.05	0.32*	0.28*	-0.01	-0.04
	TP Flat&Mod.	0.45*		0.17*	0.55*	0.18*	0.09	0.61*	0.13*	-0.26*
	TP Steep	0.30*		0.06	0.39*	0.08	0.06	0.41*	0.17	-0.13
TN	All Classes		0.39*	-0.09	0.42*	0.57*	0.76*	0.28*	0.05	0.54*
	TN Flat		0.40*	0.04	0.26*	0.64*	0.88*	0.06	0.13	0.53*
	TN Mod.		0.37*	-0.16*	0.40*	0.57*	0.75*	0.24*	0.02	0.45*
	TN Steep		0.36*	0.02	0.22*	0.72*	0.47*	0.06	0.04	0.37*

4.3.1 Chlorophyll a

A Spearman rank-order correlation analysis was conducted with TN and TP against benthic and sestonic chl-a concentrations. For benthic chl-a, a total of 192 valid samples from NMED and NRSA sites were included in the analysis. On average, NRSA benthic chl-a concentration was less than NMED chl-a (Figure 25) as NMED used a targeted richest habitat sampling, so the data were analyzed separately by source. Of all the Spearman correlations, only TP and chl-a in NMED TP High-Volcanic and all sites were significantly correlated (Table 16).

Table 16. Sample sizes (N) and Spearman rank correlation coefficients (rho) for benthic chl-a by nutrient and site class. Significant correlations (p<0.05) are marked with an asterisk (*).

	NMED		NRSA	
TP (Site Classes)	N	rho	N	rho
All sites	140	0.17*	50	-0.16
TP High-Volcanic	23	0.48*	13	-0.29
TP Flat-Moderate	90	0.16	28	-0.29
TP Steep	27	0.05	7	0.94
TN (Site Classes)				
All sites	142	0.02	50	0.09
TN Flat	26	0.04	24	0.28
TN Moderate	94	-0.04	21	0.13
TN Steep	22	-0.04	5	N/A



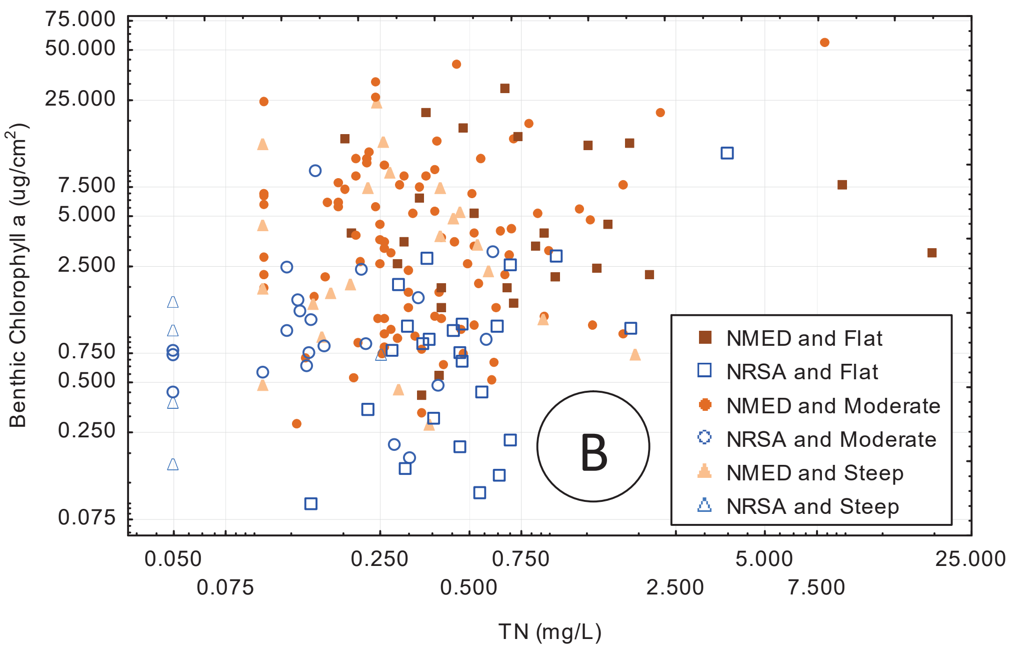


Figure 25. Benthic chl-a in relation to TP and TN, showing site classes and sources.

Positive correlations were expected between nutrients and benthic chl-a. However, a number of biotic and abiotic variables influence algal biomass accrual (Marks et al 2000). For example, hydrologic disturbances have a strong influence on algal biomass (Biggs 2000; Peterson et al. 1994). Other modifying factors that might be confounding the assumed relationship between nutrients and benthic chl-a were investigated, including canopy cover, stream flow, elevation, land use types, drainage area, latitude, longitude, conductivity, temperature, pH, and turbidity. However the dataset did not have sufficient information to account for canopy cover and stream flow influences.

Greater canopy cover (reducing light in the stream) was expected to be associated with less benthic chl-a. Canopy cover was available in all NRSA sites and few NMED sites. Mountain, foothill, and xeric sites were analyzed independently because of differences in vegetation among those regions that could affect canopy cover. Contrary to expectations, canopy cover was positively correlated to benthic chl-a in xeric and foothill sites (Spearman rho = 0.24 and 0.42, respectively), though the coefficients were not significant ($p>0.05$).

Without canopy cover data for all sites, catchment size was considered as a possible surrogate for canopy cover, assuming that larger streams would have more open canopies. Catchment size was compared to canopy cover in 38 NRSA sites. Canopy was negatively correlated to catchment size in all NRSA sites (Spearman rho = -0.16) and in mountain sites (Spearman rho = -0.29), although none of the correlation coefficients were significant ($p>0.05$). In the foothills the correlation was positive (Spearman rho = 0.20) and in the xeric areas there was almost no relationship (Spearman rho = -0.04). Because these relationships were weak, variable, and not significant, catchment size was not used as a surrogate for canopy cover.

Additional explorations of modifying factors for the relationship between nutrients and benthic chl-a are described here and detailed in Appendix H. In multiple regression and random forest analyses, benthic chl-a was related to conductivity, TP, elevation, drainage area, latitude, longitude, and pH. Of the variables considered (lacking hydrology and groundwater information), conductivity was consistently the most important predictor of benthic chl-a. Conductivity could be a stressor related to intensive land uses and not a valid classification variable. However, it might be related to natural geological and groundwater inputs. Relationships between benthic chl-a, conductivity, and covariates of conductivity (catchment size, land use types, elevation, latitude and longitude) were further investigated. Hydrologic conditions and evidence of groundwater influences were not available in enough sites to be used in the analysis.

It was determined that conductivity was positively related to catchment size and agricultural uses, though the agricultural and urban uses never amounted to any large percentages of the catchments. Elevation, percent forest cover, and longitude were more important predictors of conductivity than land uses. Conductivity was lowest in small forested catchments. Conductivity might be related to hydrology also, but flow data were sparse and were not tested. Although conductivity appears to have natural sources, when it was accounted for as a natural factor affecting benthic chl-a concentrations, remaining relationships between nutrients and benthic chl-a were not significant. Variability of chl-a within sites also affects the ability to explain benthic chl-a in terms of site characteristics.

Water Column Chl-a

Correlations between nutrients and water column (sestonic) chl-a were positive in the NRSA data set for data pooled across site classes (Table 17, Figure 26). Within the site classes, significant correlations were found in data subsets with more than 15 samples. In a random forest analysis, four variables were more important than TP but less important than TN in classifying sestonic chl-a values, including elevation (Figure 27), conductivity, turbidity, and drainage area. Sestonic chl-a was not significantly correlated with benthic chl-a (Spearman rho = 0.15, p > 0.05) or with instantaneous DO (Spearman rho = -0.16, p > 0.05). Sestonic chl-a data were not available for the NMED sites. High chl-a values (> 10ug/L) were noted for four sites (Table 18).

Table 17. Sample sizes (N) and Spearman rank correlation coefficients (rho) for sestonic chl-a by nutrient and site class. Marked correlations were significant (p<0.05).

TP	N	rho	TN	N	rho
All Sites	44	0.36*	All Sites	46	0.69*
TP High-Volcanic	10	-0.02	TN Flat	23	0.66*
TP Flat-Moderate	27	0.55*	TN Moderate	18	0.62*
TP Steep	7	0.77	TN Steep	5	0.00

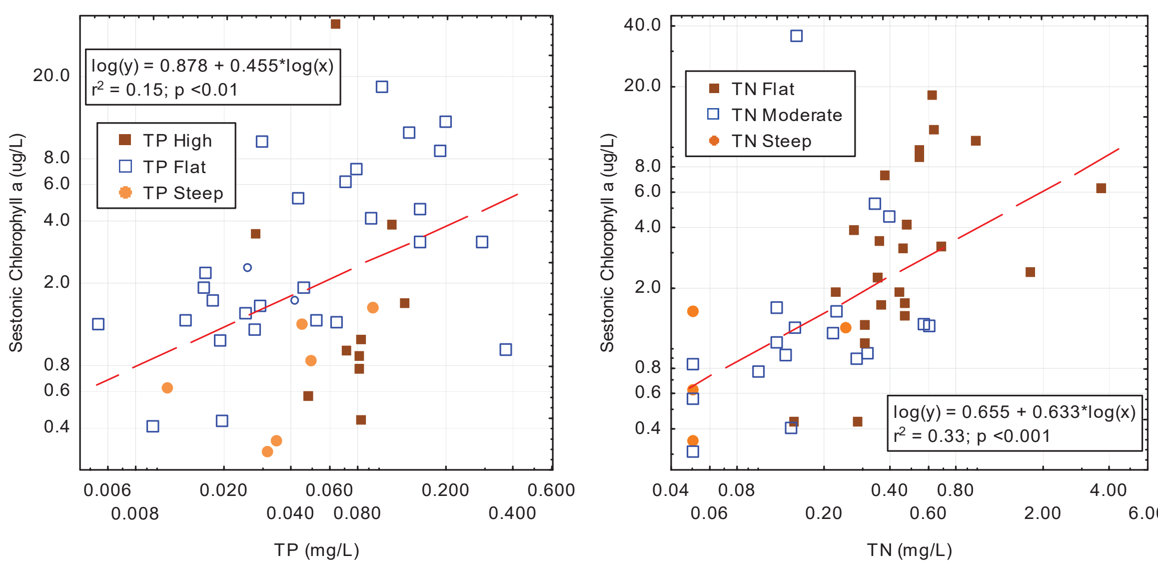


Figure 26. Sestonic chl-a in relation to TP and TN, showing site classes.

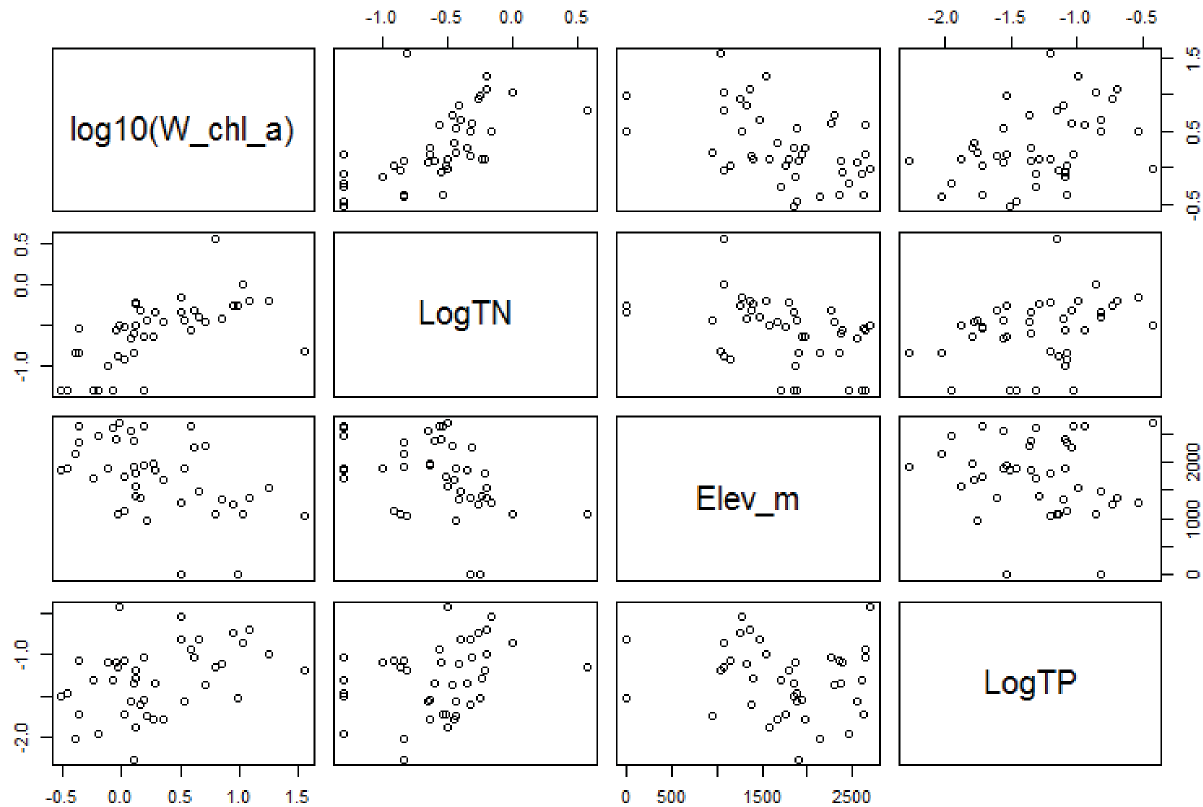


Figure 27. Relationships between sestonic chl-a and TN, elevation, and TP, using NRSA data.

Table 18. Sites with sestonic chl-a > 10ug/L.

Station ID	Stream Name	Sample Date	Chl-a (ug/L)
FW08AZ139	San Francisco River, AZ	5/7/2009	35.64
FW08NM025	Gallinas River, NM	8/2/2008	17.904
FW08NM013	Conchas River, NM	7/31/2008	12.08
FW08CO020	Wolf Creek, CO	7/29/2008	10.624

↑ Nutrients = ↑ Chlorophyll

In the correlation analyses, a positive relationship was found between benthic chl-and TP in the NMED sites. Relationships were not significant (p>0.05) for TN in NMED sites nor for TP or TN in the NRSA sites. Multiple records were at the minimum detection limit for TP. Benthic chl-a was also related to conductivity, which appears to be confounding the nutrient relationships, but once accounted for, the nutrient relationships were not evident. A positive relationship between sestonic chl-a and both TP and TN was apparent.

4.3.2 Dissolved Oxygen

A Spearman rank correlation analysis was conducted with TN, TP, and benthic chl-a against diel DO statistics. There were four record sets with minimum DO greater than 10 mg/L that were removed as outliers. Production (Pmax4hr) and respiration (Rmax4hr) were negatively correlated to each other (Spearman rho = -0.92, $p<0.05$) as were gross primary production (GPP), and ecosystem respiration (ER) (Spearman rho = -0.55, $p<0.05$). TP was positively correlated with productivity measures and maximum daily change in DO (DeltaDO) (Table 19). Both TP and TN were negatively correlated with minimum DO (rho = -0.18, $p<0.05$). TN was also negatively correlated with Rmax4hr and positively correlated with Pmax4hr and DeltaDO ($p=0.06$). Benthic chl-a was positively correlated to DeltaDO, Pmax4hr (Figure 28), and ER. The correlation with ER was expected to be negative, as it was with Rmax4hr. The bi-plots show weak positive relationships between the nutrients and DeltaDO and Pmax4hr (Figures 29 and 30) and weak negative relationships between nutrients and minimum DO (Figure 31).

Table 19. Spearman correlation coefficients for TN, TP, and benthic chl-a versus diel DO statistics; minimum DO (DOmin), maximum daily DO change (Delta DO), 4 hour maximum production (Pmax4hr), 4 hour maximum respiration (Rmax4hr), gross primary production (GPP), and ecosystem respiration (ER). Asterisk (*) denotes significant correlations ($p<0.05$).

	DO_min	DeltaDO	Pmax4hr	Rmax4hr	GPP	ER
TN	-0.18*	0.17	0.17	-0.19*	0.05	0.11
TP	-0.18*	0.30*	0.29*	-0.31*	0.19*	0.01
Benthic chl-a	-0.11	0.28*	0.38*	-0.25*	0.09	0.38*
DO_min		-0.58*	-0.50*	0.45*	-0.31*	0.04
DeltaDO			0.91*	-0.90*	0.53*	0.12
Prod4hr				-0.92*	0.62*	0.08
Resp4hr					-0.62*	-0.06
GPP						-0.55*

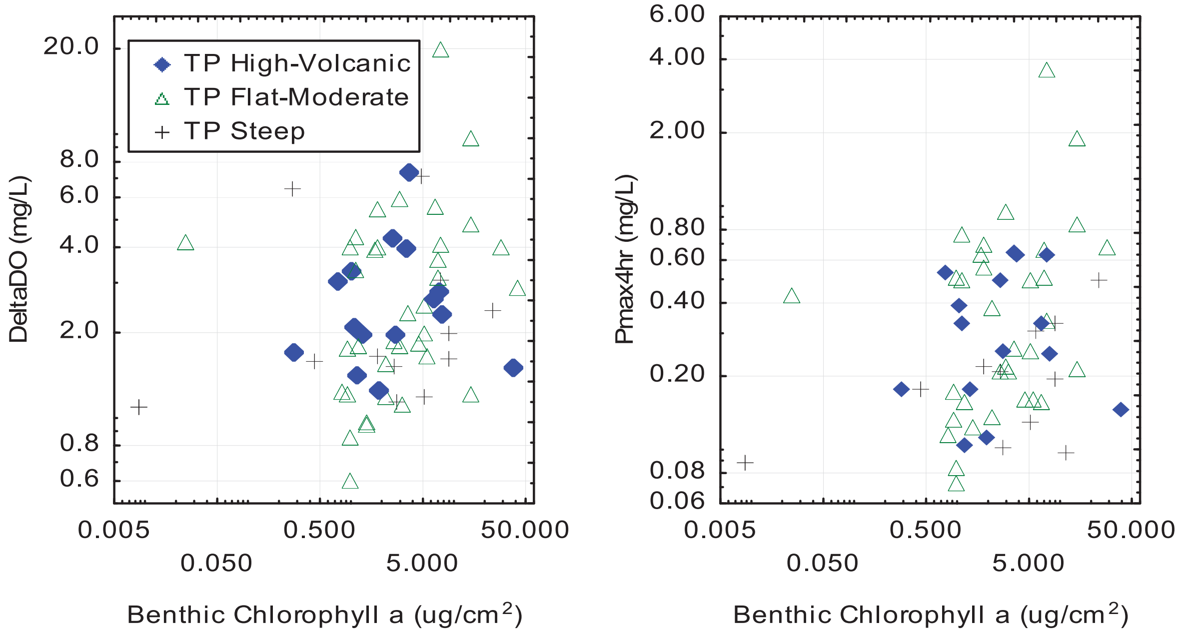


Figure 28. Benthic Chlorophyll a in relation to DeltaDO and Pmax4hr, showing TP site classes.

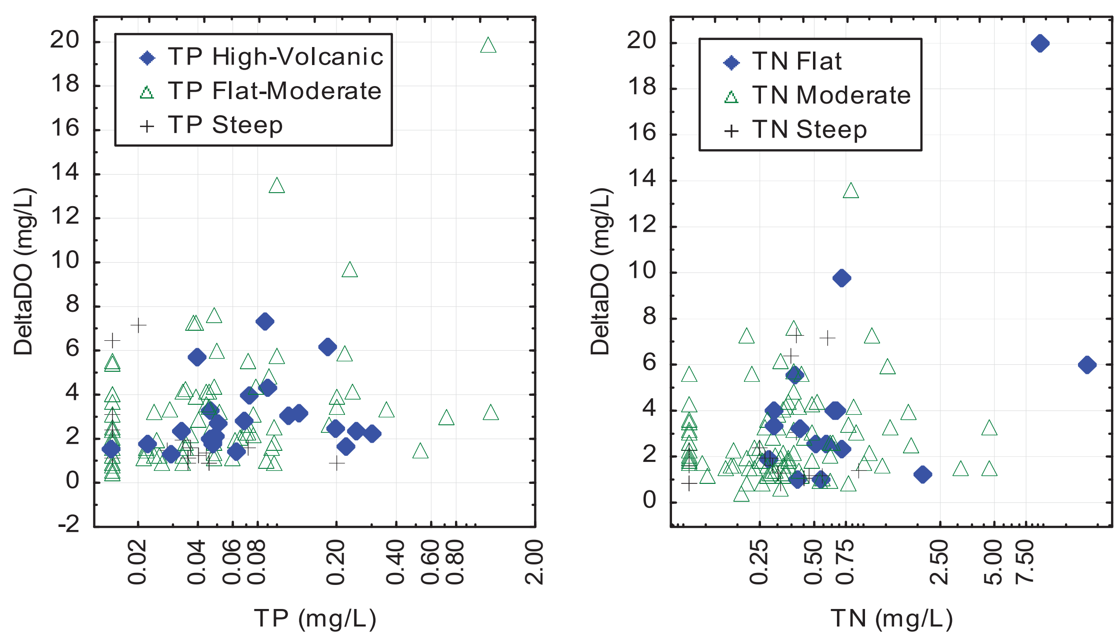


Figure 29. DeltaDO in relation to TP and TN, showing site classes.

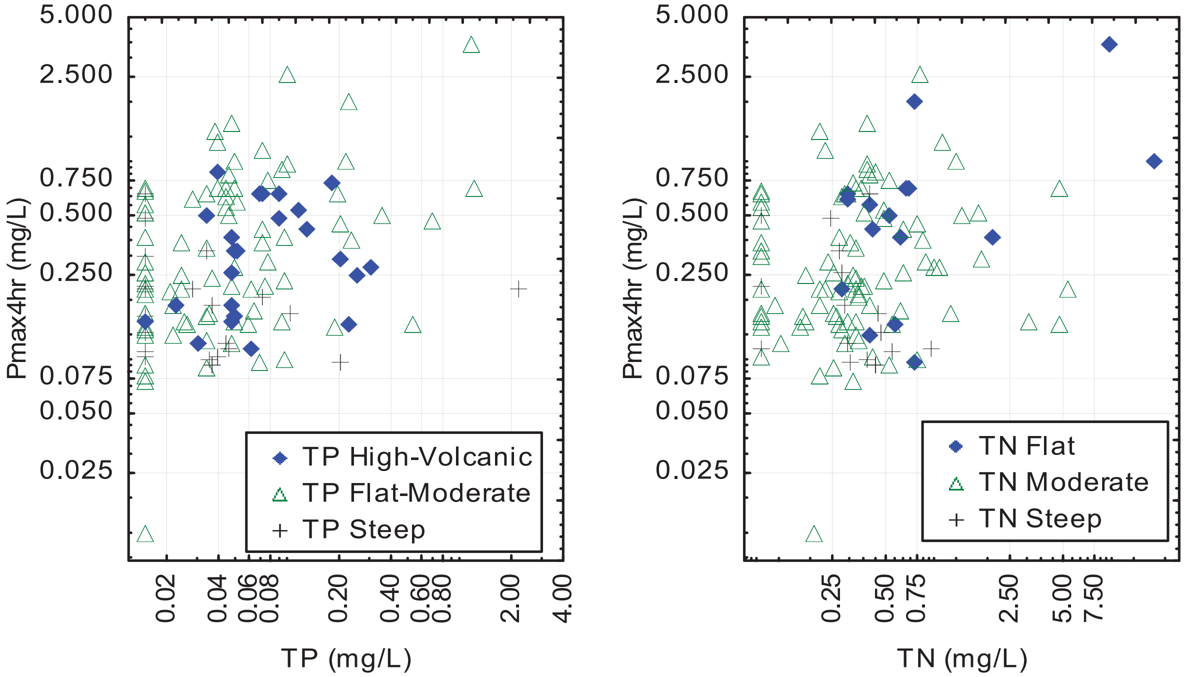


Figure 30. Pmax4hr in relation to TP and TN, showing site classes.

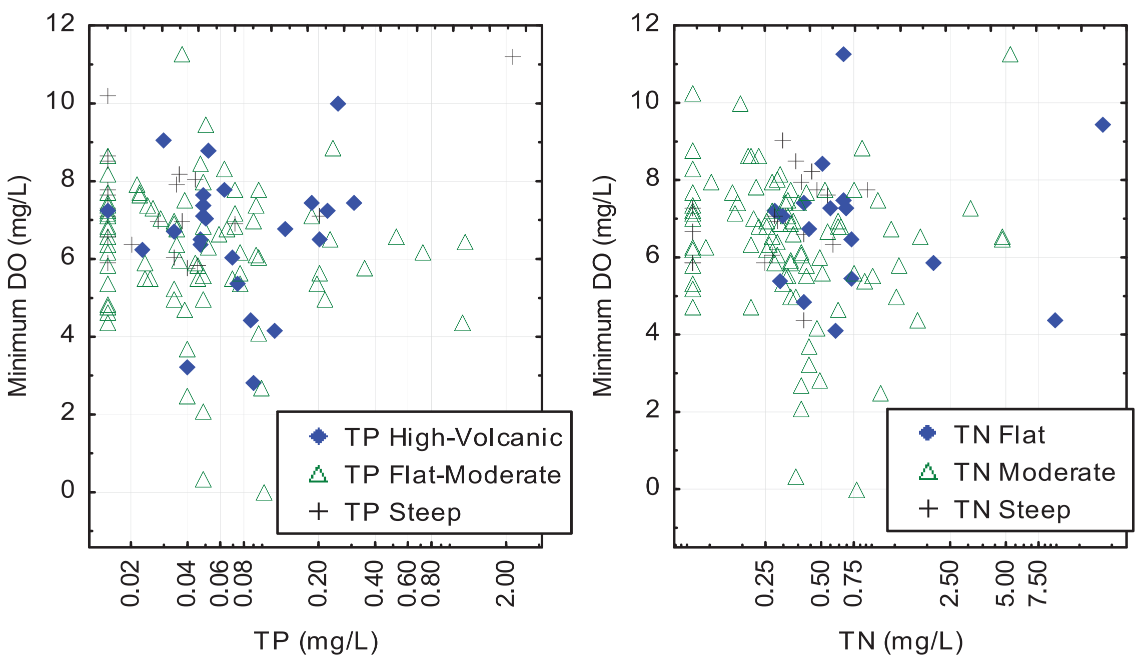


Figure 31. Minimum DO in relation to TP, showing site classes.

↑ Chlorophyll = Δ DO dynamics

Correlations of TP, TN, and benthic chl-a were positive with Delta DO and Pmax4hr and were negative with minimum DO and Rmax4hr. DeltaDO, Pmax4hr, and minimum DO were emphasized in ongoing DO analyses because of the demonstrated relationships, ability to collect continuous diel data sets using sondes and data loggers, and NMED's use of minimum DO criteria in previous nutrient assessment protocols.

4.3.3 Diatoms

A Spearman rank-order correlation analysis of nutrient values associated with diatom metrics was conducted. Sixty-eight (68) diatom metrics were correlated with TN, TP, and potential modifying factors in 151 NMED sites and 49 NRSA sites. The analysis started with NMED sites only and then addressed NRSA sites and combined data sets. The data in the analysis were limited to one sample per site when nutrient and diatom samples were collected within 30 days of each other. Details of the analysis are in Appendix I.

Diatoms appear to be more sensitive to TP than to TN, based on the number of significant correlations. The fewest significant relationships were between metrics and chl-a. Significant correlations were mostly identified in the TN Flat site class for TN and in the TP Flat-Moderate site class for TP, which had the most samples and broad range of nutrient conditions (Figure 31). This pattern was consistent for both the NMED and the NRSA data sets. TN was positively correlated to conductivity, which was also a good predictor of metric values. Conductivity (or pH or turbidity) was not factored out in the analysis of nutrient-diatom relationships. Based on significant correlations and metric types, eight responsive metrics were selected for continued analysis in stressor-response analyses (Table 20).

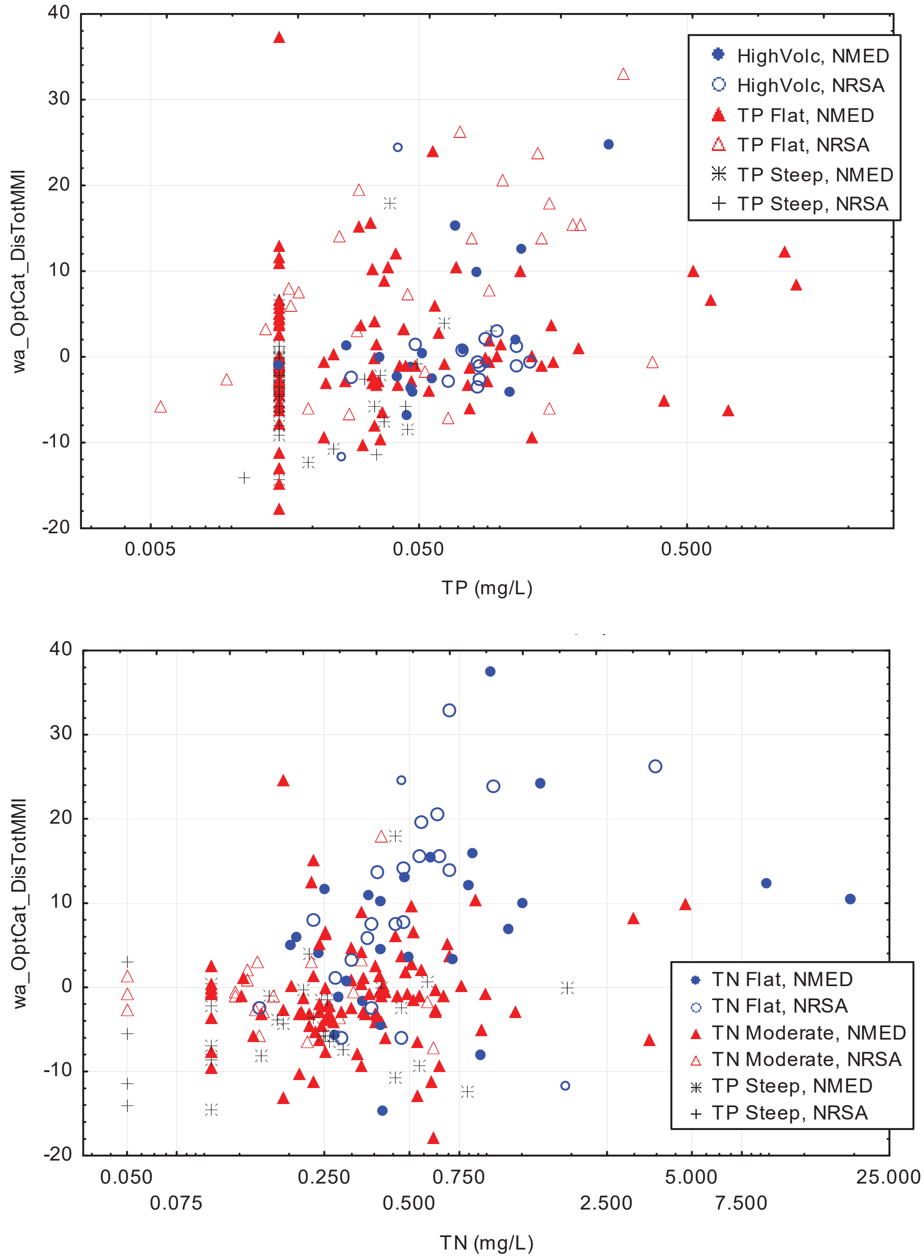


Figure 32. The diatom multi-metric index of disturbance in relation to TP and TN, showing site classes and data sources.

Table 20. Diatom metrics showing responsiveness in correlation analysis and used in stressor-response analysis.

Metric code	Metric description	Metric type
wa_OptCat_DisTotMMI	Multi-metric index of disturbance	Weighted average, general disturbance
wa_OptCat_L1DisTot	Sum of disturbances	Weighted average, general disturbance
wa_OptCat_L1Ptl	Western EMAP TP score	Weighted average, TP
wa_OptCat_LNtl	Western EMAP TN score	Weighted average, TN
wa_OptCat_NutMMI	Western EMAP multi-metric index	Weighted average, nutrients
pi_NAWQA_TN_1	% TN tolerant diatoms	Percent Individuals, TN
pi_Ptpv_TP_all_Hi	% high TP diatoms, all regions	Percent Individuals, TP
x_Shan_e	Shannon-Wiener Diversity Index	Taxa diversity

Data source (NMED or NRSA), sampling year, and sampling location (latitude and longitude) were tested as possible confounding variables that could be factored out before assessing nutrient-metric relationships. Despite indications from multiple regressions that these factors have some effects, adjustments were not made to residuals nor were the data sets further classified. Sampling year and location were related due to the sampling design, which was not a valid reason for adjustments. The NMED and NRSA data overlapped in nutrient-metric bi-plots, suggesting that keeping them separate was not necessary.

↑ Nutrients = Δ Diatom Metrics

Several diatom metrics were correlated with TN and TP. Eight metrics were selected for stressor response analysis based on responsiveness to both nutrients. Though there are apparently some modifying factors or covariates related to nutrients and diatom metrics (e.g., conductivity with TN), they were not factored out or further classified.

4.3.4 Benthic Macroinvertebrates

A Spearman rank-order correlation analysis of nutrients, diel DO, and benthic and sestonic chl-a associated with benthic macroinvertebrate metrics was conducted. The samples were limited to one per site when nutrient or chl-a and macroinvertebrate samples were collected within 30 days of each other. For diel DO statistics, the analysis was limited to DO and macroinvertebrate samples collected within the same season (within 80 days). Distributions of metric values collected with different sampling methods were overlapping in stressor-response biplots (e.g. Figure 33). Therefore, all wadeable stream samples were pooled, including multiple methods and data from NMED, WSA, and NRSA. Only early kicknet samples from NMED and low gradient samples from NRSA were eliminated. The data screening resulted in 438-440 samples for TP and TN, respectively, from 313 sites. For diel DO and benthic chl-a samples, there were 76 and 193 samples, respectively. Dissolved oxygen grab samples were collected along with macroinvertebrate samples with greater frequency than diel DO samples. However, because the variability inherent to DO over time, the DO grab data were not emphasized.

Dissolved Oxygen

The minimum DO (DO_min) and maximum 4 hour productivity (Pmax4hr) had the highest numbers of significant correlations in all sites (15-16 of 63 metrics, each) (Appendix J). Other DO statistics had fewer significant correlations (GPP: 5, ER: 4, and Rmax4hr: 10). The strongest correlations for DO_min were positive with shredder taxa, shredder percent, and Plecoptera percent (Spearman rho = 0.39, 0.34, and 0.31, respectively). Other metrics with high positive correlations (rho = 0.30) included Ephemeroptera taxa, Beck's index, and intolerant percent. Gastropod percent and the HBI were negatively correlated (rho = 0.30). EPT taxa, a familiar and responsive metric, was positively correlated and showed that responses were similar among sampling methods (Figure 32). Low DO (<5.5 mg/L) was consistently related to lower EPT taxa. Low EPT taxa were also observed with higher DO_min, possibly due to other stressors including the alteration of habitat by abundant algal growth.

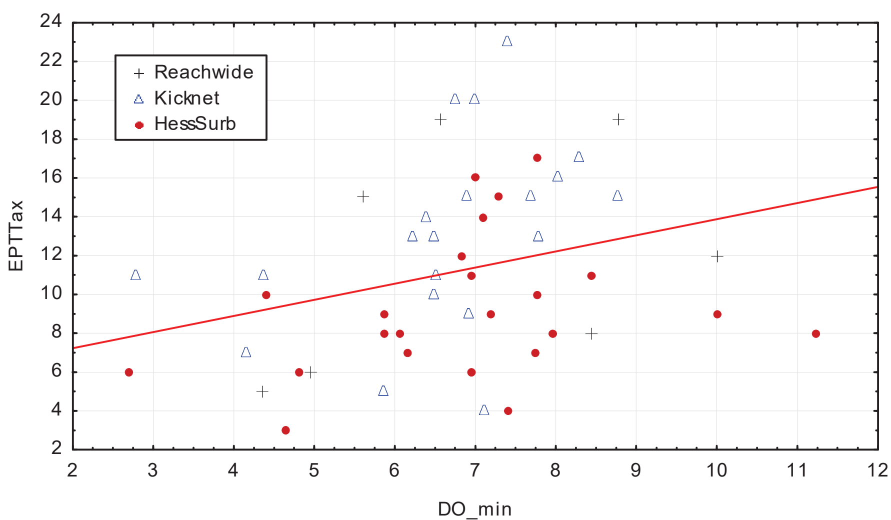


Figure 33. Relationship between EPT taxa richness and minimum DO (Spearman rho = 0.27).

With productivity (Pmax4hr), the strongest metric correlations were negative (rho = -0.37 - -0.42) with Plecoptera taxa, Plecoptera percent, Beck's index, and intolerant taxa (Appendix J). These relationships and other metric responses were similar in the TP Flat-Moderate site class.

Fewer metrics were significantly correlated to productivity in the TP High-Volcanic and TP Steep site classes.

Benthic chlorophyll

Benthic chl-a was correlated to 14 macroinvertebrate metrics in all sites. The strongest correlations were with the Shannon-Wiener diversity index and intolerant taxa (Spearman rho = 0.37 and -0.34, respectively). A positive correlation with Shannon-Wiener diversity (Figure 34) and a positive correlation with Trichoptera taxa and percent Trichoptera indicates that higher benthic chl-a increases some aspects of the macroinvertebrate assemblage diversity. Similar correlations were observed in the individual site classes, especially in the TP Flat-Moderate sites.

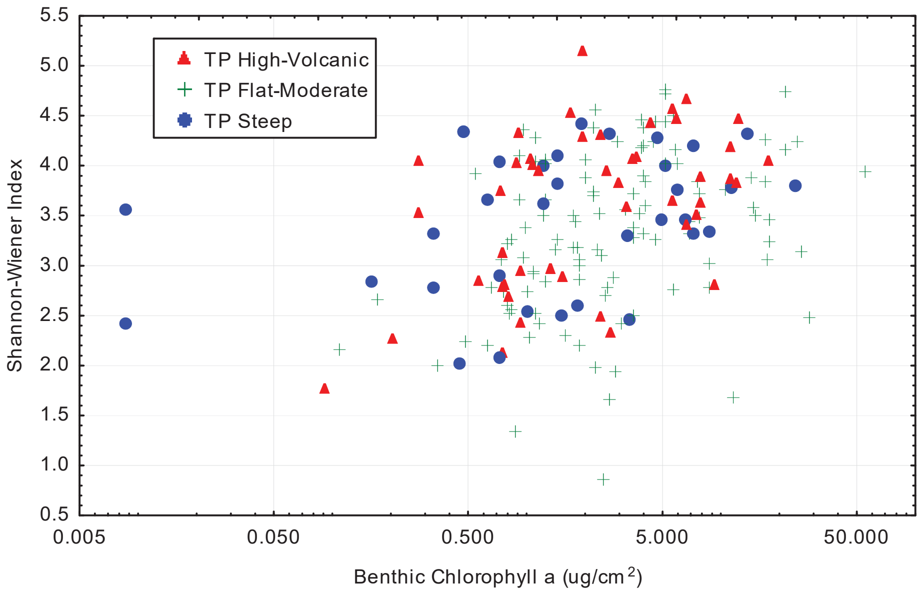


Figure 34. Metric values (Shannon-Wiener index) against benthic chl-a concentrations, marked by site class.

Sestonic chlorophyll

Sestonic chl-a in NRSA samples was negatively correlated ($p<0.05$) to six macroinvertebrate metrics, including total taxa, ETP taxa, Plecoptera taxa, intolerant taxa, percent Plecoptera, and clinger taxa. Correlation coefficients were between -0.34 and -0.43. More intolerant taxa were generally found in sites with less than 2 µg/L sestonic chl-a (Figure 35).

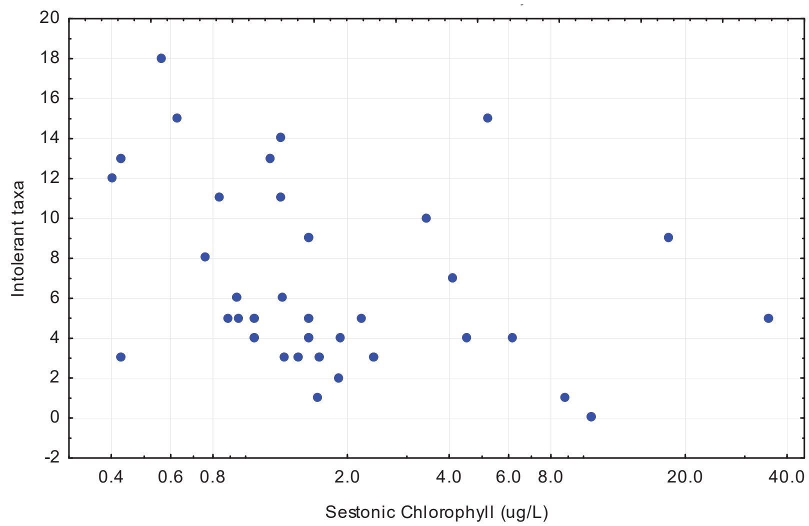


Figure 35. Metric values (intolerant taxa) against sestonic chl-a concentrations.

Nutrients

For TN, 42 of the 63 metrics were significantly ($p<0.05$) correlated with nutrient concentration in all sites (Appendix J). Fewer metrics were significantly correlated in each site class, with 28, 25, and 6 in the TN Flat, TN Moderate, and TN Steep classes, respectively. Only predator taxa and % predators were correlated in all site classes. The strongest correlations (all negative) were in the TN Flat and TN Moderate classes for the total taxa metric (Figure 36A), EPT taxa metric, Beck's Biotic Index (weighted richness of sensitive taxa, Figure 36B), and the clinger taxa metric. These metrics represent richness of sensitive or specialized organisms. In TN Steep sites, the % predators metric had one of the stronger negative correlations with TN (Figure 36C).

The relatively unresponsive metrics in the steep site class might be due to lower nutrient concentrations in general, with fewer values >0.5mg/L TN than in the TN Flat and TN Moderate site classes. Conversely, nutrient concentrations in the TN Flat site class are generally higher than in the other classes. The TN Moderate site class is the largest and shows a typical wedge-shaped plot for many metric responses to TN (Figures 36b, 36C).

For TP, fewer correlations were significant and the strength was weaker, compared to the TN correlations (Appendix J). There were 21 significant ($p<0.05$) correlations when including all site classes. Correlations varied among site classes. In the TP High-Volcanic, TP Flat-Moderate, and TP Steep classes, 17, 12, and 6 metrics were significantly correlated, respectively. Percent Coleoptera had the strongest of all correlations in the TP High-Volcanic site class (Figure 37A). Percent scrapers were negatively correlated in all site classes (Figure 37B). EPT taxa were correlated in the TP High-Volcanic and TP Flat-Moderate sites (Figure 37C). Of the few metrics correlated in the TP Steep class, Oligochaete metrics were notable increasers. Macroinvertebrates appear to be more responsive to TN than to TP. The positive correlation between TN and TP was not very strong in this data set (Spearman rho = 0.28, $p < 0.05$).

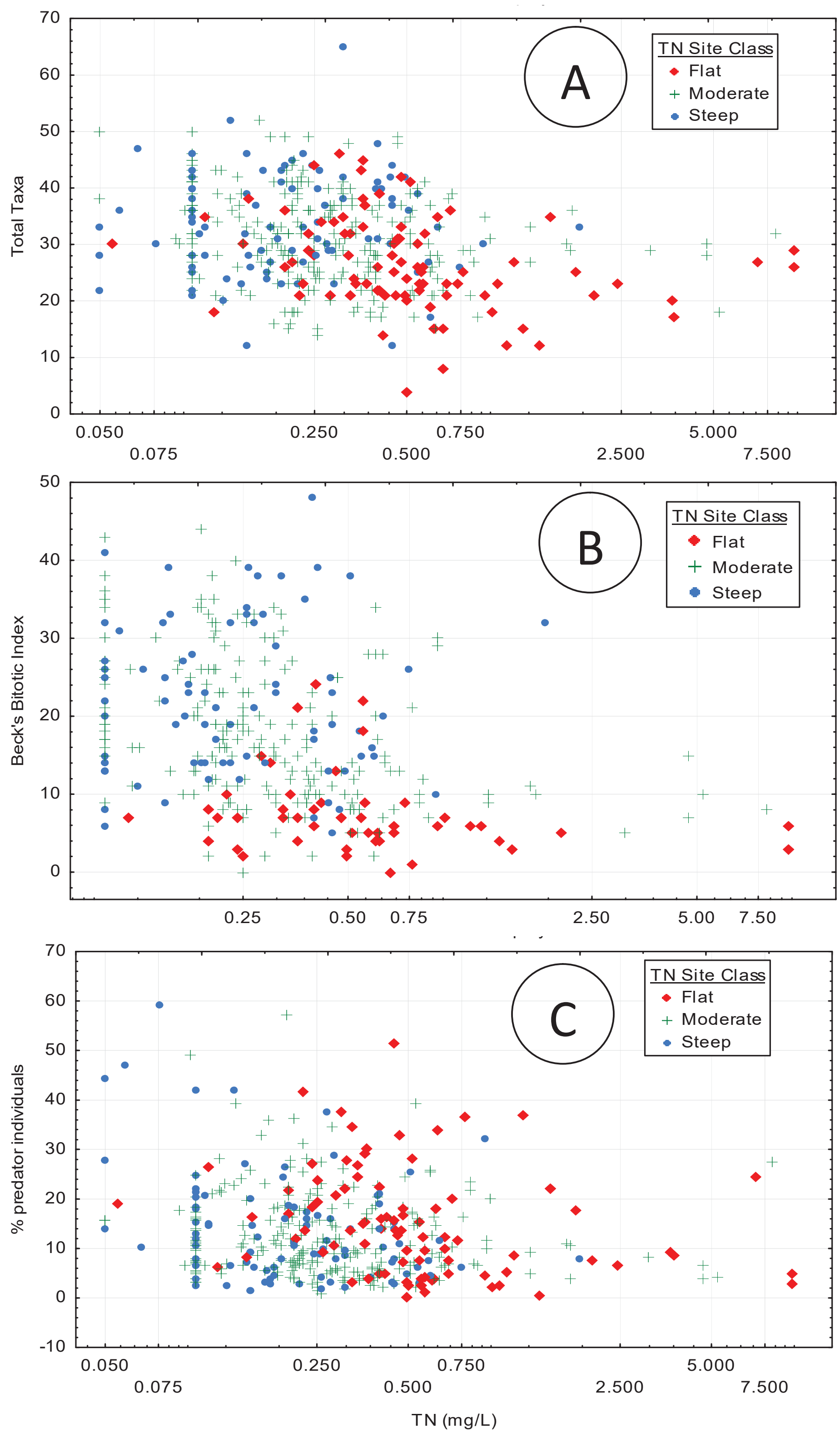


Figure 36. Metric values (A: total taxa, B: Beck's Biotic Index, and C: % predator individuals) against TN concentrations, marked by site class.

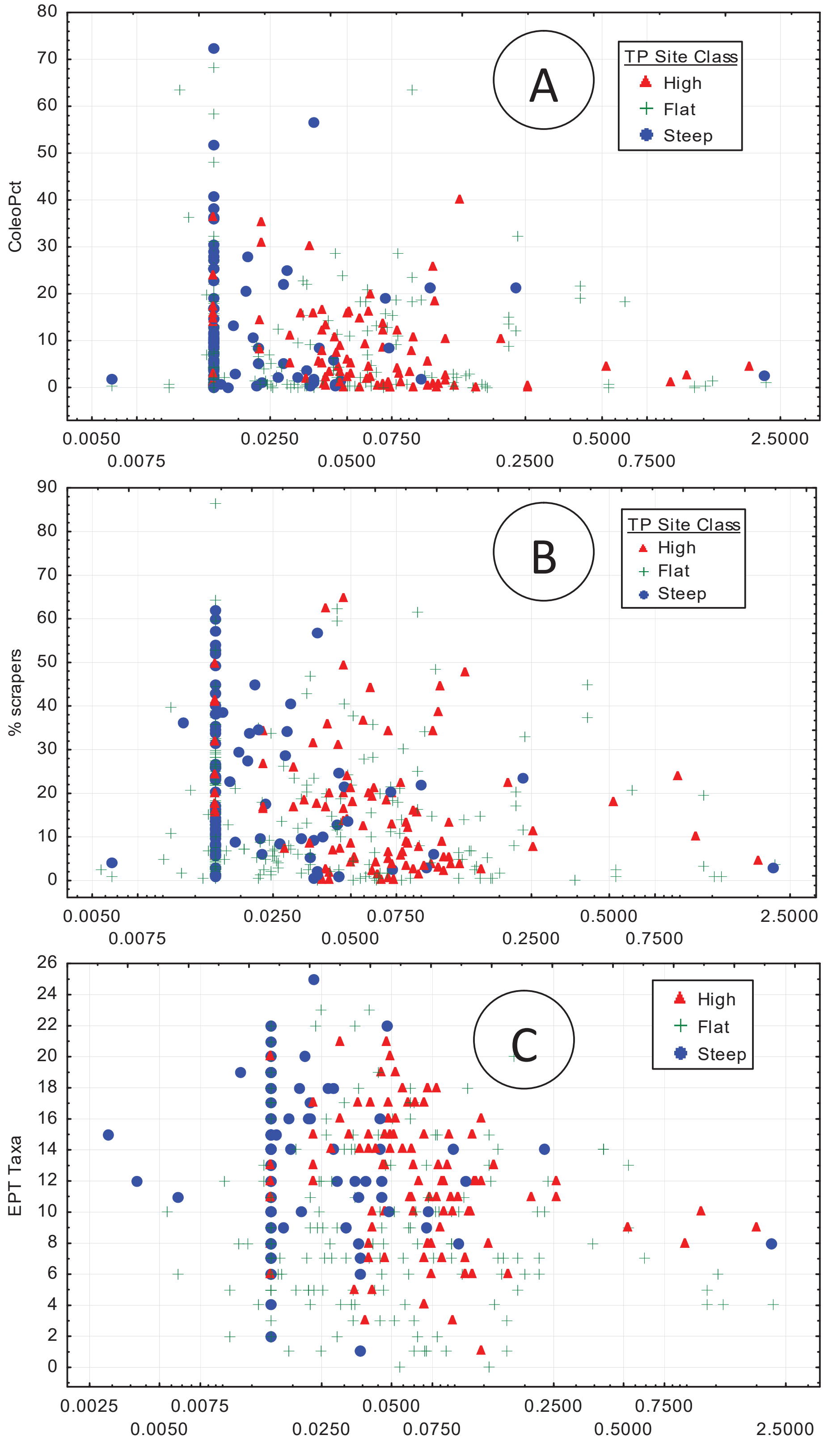


Figure 37. Metric values (A: % Coleoptera, B: % scrapers, and C: EPT taxa) against TP concentrations, marked by site class.

Several benthic macroinvertebrate metrics were related to nutrients, benthic chl-a, and DO, but only ten were selected for ongoing analyses to simplify interpretation of the stressor-response relationships. Responses for 19 candidate metrics were qualified as strongly positive, positive, negative, or strongly negative in relation to multiple measures of nutrients, benthic chl-a, and DO. The ratings were based on correlation coefficients in all sites and in the individual site classes or methods (Appendix J). Ten benthic macroinvertebrate metrics with consistent and strong correlations were identified (Table 21, bold font). These included EPT taxa, Ephemeroptera taxa, Plecoptera taxa, percent EPT, percent Plecoptera, percent non-insects, intolerant taxa, percent tolerant, shredder taxa, and percent clinger. The percent EPT metric was selected despite weak responses because it is a familiar metric that summarizes sample composition of three generally sensitive taxa groups.

Table 21. Qualitative response trends for macroinvertebrate metrics to nutrients, benthic chl-a, and DO. The trends of responses were negative (Neg) or positive (Pos). Stronger relationships (more significant correlations in site classes) are shown in bold type.

		TN	TP	Chl-a	DO[a]	Overall
Richness	Total Taxa	Neg	Neg	Pos	Mix	Mix
	1 EPT Taxa	**Neg**	**Neg**	Mix	**Neg**	Neg
	2 Ephemeroptera Taxa	**Neg**	**Neg**	Neg	**Neg**	**Neg**
	3 Plecoptera Taxa	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**
	Trichoptera Taxa	**Neg**	**Neg**	Pos	Mix	Mix
	Shannon-Winer Index	**Neg**	**Neg**	**Pos**	Mix	Mix
Composition	**4 EPT percent**	Neg	Mix	Mix	Mix	Neg/Mix
	Ephem percent	Mix	Mix	Mix	Mix	Mix
	5 Pleco percent	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**
	Trich percent	**Neg**	**Neg**	**Pos**	Mix	Mix
	6 NonIn percent	**Pos**	Pos	Pos	Pos	**Pos**
Tolerance	**7 Intolerant Taxa**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**
	8 Toler percent	**Pos**	Pos	Pos	**Pos**	**Pos**
Feeding Group	Cllct percent	**Pos**	**Pos**	**Neg**	Mix	Mix
	Scrap percent	**Neg**	**Neg**	**Pos**	Mix	Mix
	Shred percent	**Neg**	Neg	**Neg**	**Neg**	**Neg**
	9 Shredder Taxa	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**
Habit	Brrwr percent	**Pos**	**Pos**	**Neg**	Pos	Mix
	10 Clngr percent	**Neg**	**Neg**	Mix	Mix	Neg

a: The DO measures characterized in the qualitative correlations were Pmax4hr and GPP, which gave opposite responses compared to Rmax4hr and minimum DO.

The following series of plots illustrates several linkages between nutrients and benthic macroinvertebrates, focusing on the intolerant taxa metric, which was negatively related to nutrients and chl-a and positively related to minimum DO (see Appendix J). The immediate stressor on intolerant taxa was assumed to be DO. Field measured DO might have been taken at any time of day, adding variability to the measure, so the analysis emphasized minimum DO and Pmax4hr. As minimum diel DO decreased, especially below 6 mg/L, intolerant taxa decreased in all site classes (Figure 38).

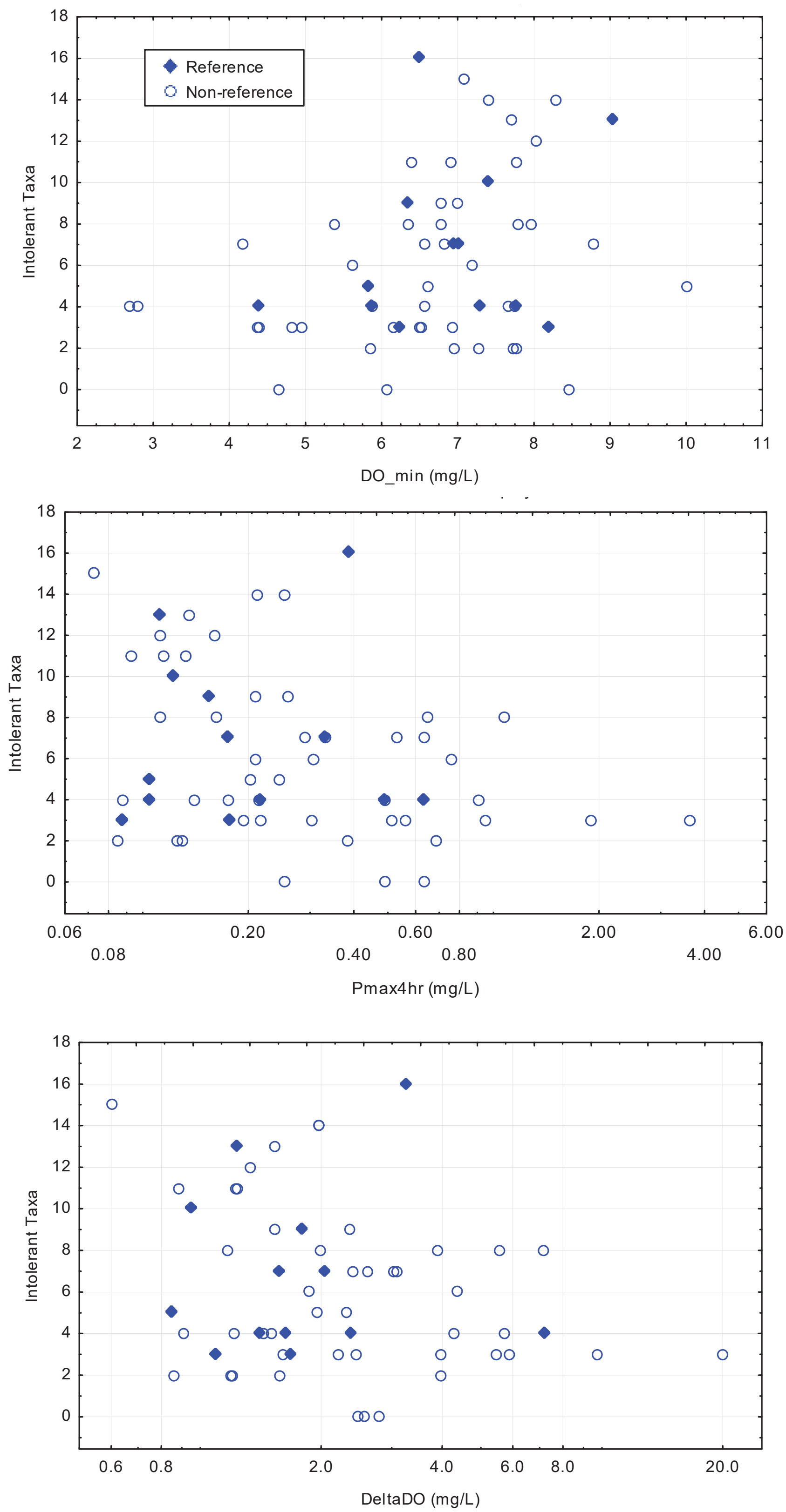


Figure 38. Intolerant taxa metric versus minimum DO, Pmax4hr, and DeltaDO, marked by reference status.

Intolerant taxa generally decline with increasing benthic chl-a, though there are also low numbers of intolerant taxa with low chl-a (Figure 39). Macroinvertebrates generally have an optimal preference across stressor gradients with unimodal relationships, so low numbers on either side of the optimum is expected (USEPA 2006). The low chl-a, low intolerant taxa sites might be either oligotrophic (supporting few taxa because of generally low production) or toxic (containing an unmeasured toxicant though nutrients are low). Minimum DO and chl-a are not strongly related (Figure 40). However, the relationships of chl-a with Pmax4hr and DeltaDO are somewhat stronger, supporting the causal linkage between chl-a and DO.

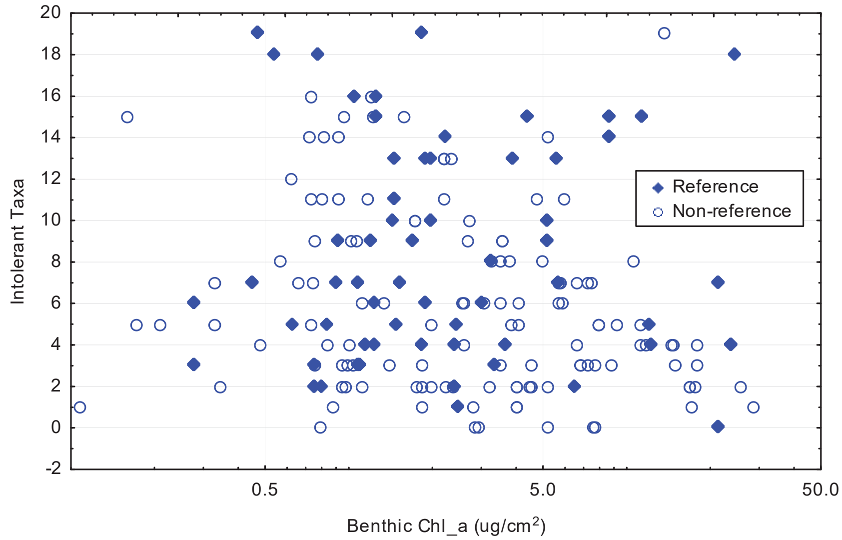


Figure 39. Intolerant taxa metric versus benthic chl-a, marked by reference status.

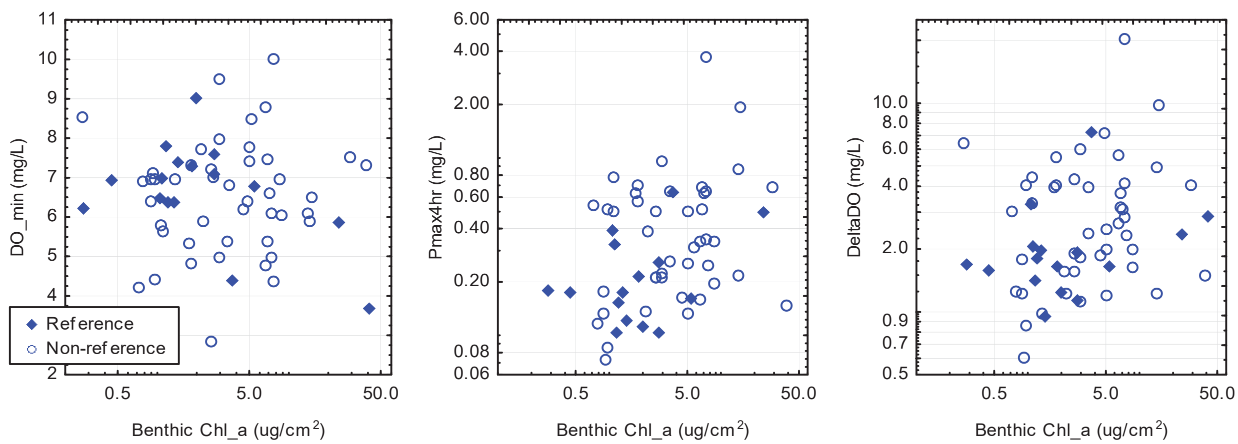


Figure 40. Relationship between benthic chl-a and DO measures; minimum DO, Pmax4hr, and DeltaDO, marked by reference status.

The final link to nutrients is seen in the relationships between TP, TN, and benthic chl-a (Figure 41). The relationships with chl-a were variable, so the indirect relationships between nutrients and DO were explored (see Figures 29-31), showing increased DeltaDO and Pmax4hr with increasing nutrients, and decreasing minimum DO. In another indirect relationship, nutrients were compared to the intolerant macroinvertebrate metric (Figure 42). Fewer intolerant taxa were associated with increasing nutrient concentrations. Macroinvertebrate metrics and nutrients were examined in stressor-response analyses because, though the effects are indirect, the relationships were relatively clear. The intermediate stressors (chl-a and DO) showed trends that support the causal model, though they were variable. The chl-a and diel DO data sets were smaller in comparison to the macroinvertebrate metric data set, which might affect the apparent strength of relationships with nutrients.

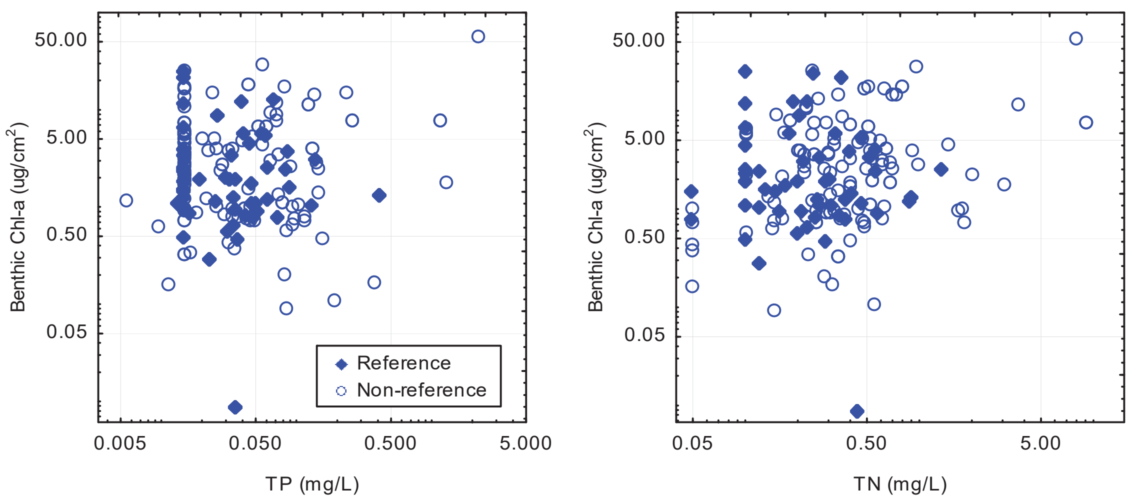


Figure 41. Benthic chl-a versus TP and TN, marked by reference status**.**

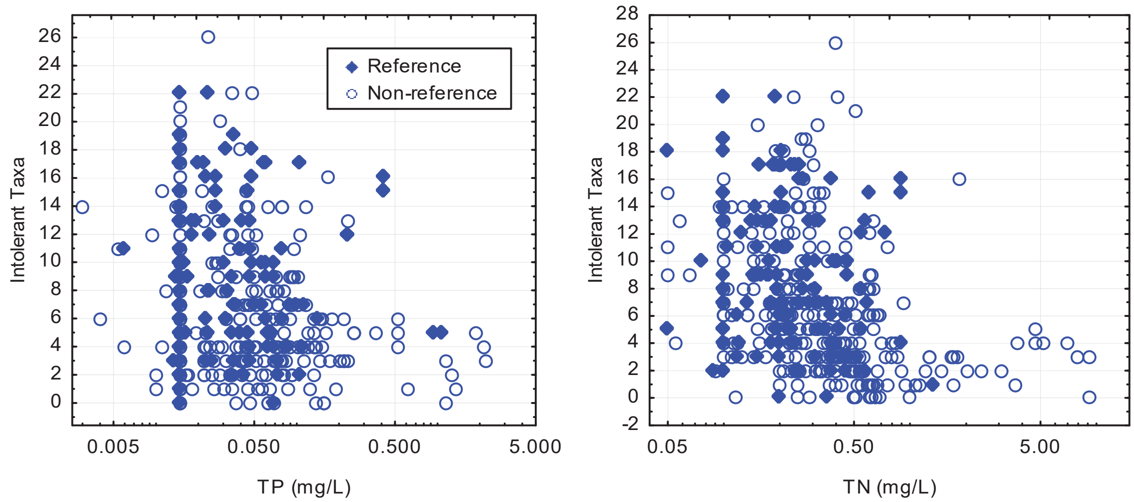


Figure 42. Intolerant macroinvertebrate taxa versus TP and TN, marked by reference status.

In a final comparison, intolerant taxa were found to be sensitive to other stressors, such as conductivity measured within 30 days of the macroinvertebrate sample (Figure 43). Conductivity was correlated with nutrients, especially TN (see Table 15). The nutrient-macroinvertebrate relationship was not adjusted for such co-occurring factors.

Figure 43. Relationship between intolerant taxa and conductivity (EC), marked by reference status.

↓ Minimum DO = Δ Macroinvertebrate Metrics

As with diatom metrics, the most responsive macroinvertebrate metrics were selected for continuing analysis of stressor-response effects. These included 10 metrics that are commonly used in bioassessments. In the conceptual model, macroinvertebrates respond directly to minimum DO conditions, which are related to chl-a. Other measures of DO (Pmax4hr and DeltaDO) were also related to macroinvertebrate metrics, though these were not specified in the conceptual model. However, in these analyses, the strongest relationship was directly between nutrients and macroinvertebrate metrics (bypassing DO or chl-a). This might be due to a larger data set for nutrients relative to datasets for DO or chl-a. Different macroinvertebrate sampling methods were indistinct in biplots of stressors and metrics. Therefore, data from multiple sampling methods were pooled in stressor-response analyses.

4.4 Regression Interpolation

Ten macroinvertebrate and eight diatom metrics were regressed against TN and TP. The regressions included all nutrient and biological samples that were taken from the same site within a 30 day window. Data from all site classes were used to derive the regression equations because this assured a complete nutrient gradient on the x-axis. Regressions within the individual site classes would have resulted in shorter stressor gradients. In some classes, the gradients would not include enough high nutrient concentrations to show significant relationships.

The scatter plots of metrics on nutrients were often wedge-shaped, with the heel of the wedge containing samples with poor response metrics though the nutrient stressors were low (Figure 44). These samples might have poor metrics because of other stressors, such as sediment and temperature. The entire data set was used to derive the regression equations because sufficient information for removing samples with non-nutrient stressors was not available. This resulted in a mean regression slope that is not as steep as the effective slope observed as a regression of the upper quantiles of the data. Shallow slopes of the regression equations result in large changes in interpolated nutrient values for each incremental change of the reference metric value.

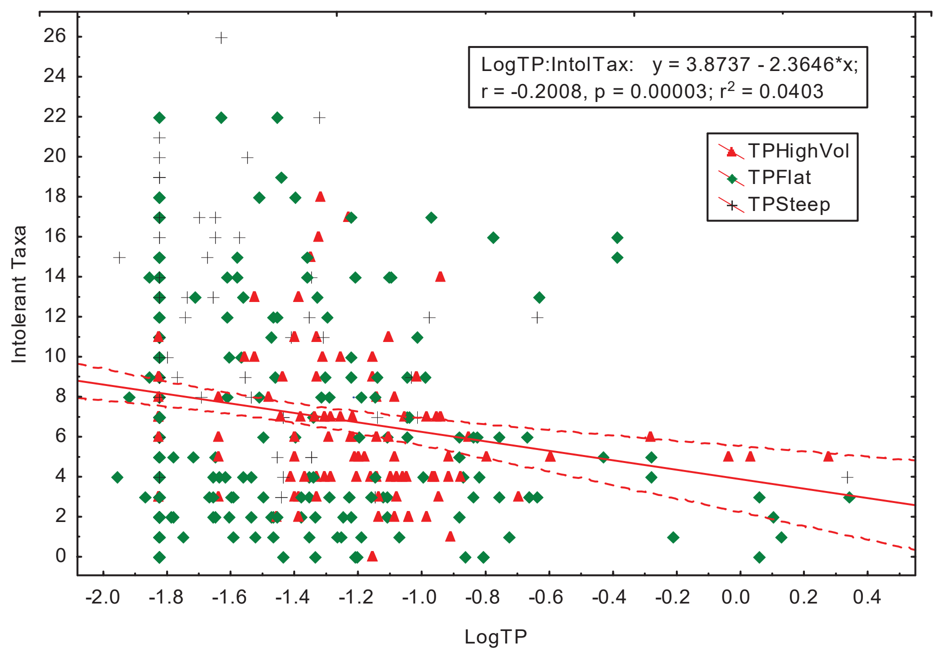


Figure 44. Relationship between TP and intolerant macroinvertebrate taxa, showing the regression equation with sites marked by site class.

The reference 25th (or 75th) quantile of metric values in each nutrient site class were interpolated to a nutrient value on the x-axis. These quartiles of reference observations were selected as the critical values to represent reference expectations. Metric quartiles for the TN Flat class were substantially different than those in the other classes. For example, the EPT taxa quartile was 10 and 12 in the TN Steep and TN Moderate classes, but was only 5 in the TN Flat class. This led to invalid interpolated values for many metrics. The high-small multi-metric macroinvertebrate condition index was not used, though it had an associated threshold of impact, because it could only be applied in a limited number of sites.

The nutrient values associated with reference metric quartiles were interpolated by substituting the critical metric value as y in the equation and solving for x (Appendix K). The results were only considered as candidate nutrient thresholds if the regression equation was significant ($p<0.05$) and the interpolated nutrient value was within the range of observed values. Any values extrapolated beyond the observed range (excluding extreme values) in each site class were disregarded.

The resulting valid candidate thresholds ranged from 0.13 to 3.26 mg/L for TN and from 0.003 to 1.74 mg/L for TP (Table 22). For TN, all metrics except the Shannon-Wiener index for diatoms were significantly related to nutrients in at least one site class. Median candidate threshold values in the TN Moderate and TN Steep site classes were close to the values derived from the 90th quantile of the reference distributions, though comparatively lower in the TN Moderate sites and higher in the TN Steep sites.

For TP, six of the ten macroinvertebrate metrics were significantly related to nutrient concentrations using regression interpolation (see Appendix K). All of the TP values interpolated from macroinvertebrate metrics were higher than the 90th quantiles of the reference distributions and the interpolated values that were retained in the analysis were at the high ends of the observed ranges. All diatom metrics except the Shannon-Wiener index were significantly related to TP concentrations. Interpolated TP values from diatom metrics were generally lower than the 90th quantiles of the reference distributions in the TP High-Volcanic and TP Steep site classes, and comparatively higher in the TP Flat-Moderate site class.

Regression interpolation of TN and TP from critical minimum DO of 5 or 6 mg/L did not yield valid results. The regression equations were not significant ($p > 0.10$) and the interpolated values were orders of magnitude greater than the observed maximum nutrient concentrations. Critical values for maximum productivity (log transformed) were derived from their regression with critical minimum DO values ($r^2 = 0.26$, $p<0.001$). Interpolated nutrient concentrations related to 6 mg/L minimum DO (Pmax4hr = 0.50 mg/L) were 4.4 mg/L TN and 0.98 mg/L TP. These results were high compared to the reference 90th quantile values. Interpolated nutrient concentrations related to 5 mg/L minimum DO were orders of magnitude greater than the observed nutrient concentrations and were therefore disregarded. This analysis was conducted on a statewide basis, not within site classes.

Table 22. Candidate thresholds derived from regression interpolations on selected macroinvertebrate and diatom metrics. Values in gray font were not valid because they did not have significant regression equations or were outside of the observed range of values in the site classes.

	TN (mg/L)			TP (mg/L)		
	TN Flat	TN Moderate	TN Steep	TP High-Volcanic	TP Flat-Moderate	TP Steep
EPTTax	3.70	0.18	0.43	0.11	3.39	0.11
EphemTax	2.27	0.62	0.62	1.00	211	1.00
PlecoTax	3.26	3.26	3.26	1.61	1.61	0.15
IntolTax	3.48	0.53	0.85	0.88	6.22	0.33
Toler percent	501	0.29	0.33	0.22	3.11	0.017
EPT percent	398	0.13	21.47	59102	31.91	491758
Pleco percent	0.49	0.49	0.49	1.74	1.74	0.80
NonIn percent	28.33	0.21	0.37	1.46	0.281	0.003
ShredTax	2.28	2.28	0.64	56.47	56.47	0.60
Clngr percent	108	0.88	2.44	13.22	5.01	0.50
BMI Medians	2.28	0.49	0.46	0.11	1.61	0.11
wa_OptCat_DisTotMMI	10.35	0.36	0.19	0.042	0.168	0.028
wa_OptCat_L1DisTot	18.26	0.30	0.19	0.024	0.358	0.027
wa_OptCat_L1Ptl	7.45	0.43	0.29	0.068	0.145	0.029
wa_OptCat_LNtl	10.49	0.33	0.18	0.057	0.311	0.054
wa_OptCat_NutMMI	9.26	0.32	0.23	0.047	0.193	0.025
pi_NAWQA_TN_1	1.28	4.36	5.32	0.457	0.129	0.010
pi_Ptpv_TP_all_Hi	7.98	0.25	0.69	0.083	0.152	0.011
x_Shan_e	2.26	16.63	161700	9.272	7.272	0.012
Diatom Medians	1.28	0.33	0.21	0.052	0.168	0.026
Median of all valid interpolated values	2.27	0.33	0.35	0.063	0.237	0.025
Reference 90th quantile	0.69	0.42	0.30	0.105	0.071	0.054
Maximum in site class	3.44	2.63	0.75	0.22	1.82	0.12

For the regression interpolation of DO statistics on nutrient concentrations, both nutrient concentrations and DO stats were log transformed. The regression equations were calculated with all sites classes combined. Regression equations in the three individual TP site classes resulted in non-significant regressions in the TP High-Volcanic and TP Steep classes. However, in the TP Steep class, the relationships between TP and both Delta DO and Pmax4hr were negative and significant ($p<0.05$). The negative relationships were only seen in the TP Steep site class. Equations for the TP Flat-Moderate class were similar to those in all sites, so we emphasized results from equations for all site classes combined (Table 23). Regression interpolation in the

TP High-Volcanic and TP Steep site classes were at the extreme high and low (respectively) ends of the range of observed values.

Table 23. Candidate thresholds for DO statistics derived from regression interpolations on reference 90th quantile nutrient concentrations. Values in gray font were not valid because they were at the extremes of the range of values.

	Delta DO			Pmax4hr		
	TP High-Volcanic	TP Flat-Moderate	TP Steep	TP High-Volcanic	TP Flat-Moderate	TP Steep
TN	16.39	3.34	1.13	3.23	0.47	0.13
TP	12.63	4.06	0.92	2.36	0.56	0.08

4.5 Change-point Analysis

The change-point analysis always identifies a change-point, but that change-point can be subject to the shape of the response curve. If there was an emphasis on either least disturbed or highly disturbed sites in site selection, the concentration of samples could affect the location of the change-point. Confidence intervals were calculated for each change-point to illustrate the possible ranges of change-points. All of the change-point graphs are presented in Appendix L. The decisions to accept the change-points as valid candidate thresholds were based on valuation of the 95th quantile regression line, the relative size of the confidence interval around the change-point, and coincidence of an appropriate slope in the LOWESS regression line at the change-point.

Change-points were identified for both TN and TP from 10 macroinvertebrate metrics, 8 diatom metrics, and 2 DO measures (Table 24). The ranges of valid change-points were fairly narrow for each nutrient and site class (at most 1.24 mg/L for TN and 0.08 mg/L for TP). For TN, median candidate thresholds were greatest in the TN Flat site class and least in the TN Steep site class. Likewise for TP, TP Steep sites had lower median change-points and increasing change-point medians were in the TP Flat-Moderate and TP high-Volcanic classes. The medians of all valid change-points was lower than the 90th quantile of the reference distributions for each site class and nutrient (Table 24). The EPT taxa, Plecoptera taxa, weighted average disturbance (wa_OptCat_L1DisTot), and weighted average nitrogen preference (wa_OptCat_LNtl) metrics had the most valid change-points associated with them.

For TN, the change-points derived from DO measures were similar to the 90th quantile of reference values, even in the TN Flat site class, where the change-points were not valid. All change-points for macroinvertebrates and diatoms were valid in the TN Flat site class and most of them were lower than the reference 90th quantile value. Valid TN change-points were variable in the TN Moderate and TN Steep site classes, with only one valid change-point derived from macroinvertebrate metrics in the TN Steep class.

Table 24. Change-points (CP) as candidate thresholds from selected benthic macroinvertebrate (BMI), diatom and dissolved oxygen (DO) metrics. Values in gray font did not pass the tests for valid change-points.

Metric	TN (mg/L) TN Flat	TN Moderate	TN Steep	TP (mg/L) TP High-Vol	TP Flat-Moderate	TP Steep
EPTTax	0.49	0.25	0.42	0.067	0.044	0.030
EphemTax	0.49	0.22	0.28	0.058	0.044	0.030
PlecoTax	0.56	0.33	0.25	0.063	0.041	0.027
IntolTax	0.48	0.29	0.39	0.061	0.051	0.029
Toler percent	0.66	0.40	0.26	0.083	0.052	0.041
EPT percent	0.97	0.36	0.22	0.047	0.014	0.029
Pleco percent	0.35	0.33	0.14	0.114	0.044	0.027
NonIn percent	0.72	1.26	0.23	0.083	0.014	0.018
ShredTax	0.48	0.25	0.23	0.047	0.151	0.017
Clngr percent	1.09	0.49	0.28	0.122	0.051	0.022
Median CP BMI	0.53	0.31	0.28	0.063	0.044	0.029
wa_OptCat_DisTotMMI	0.48	0.52	0.16	0.068	0.056	0.035
wa_OptCat_L1DisTot	0.50	0.38	0.26	0.068	0.066	0.034
wa_OptCat_L1Ptl	0.48	0.52	0.13	0.066	0.032	0.036
wa_OptCat_LNtl	0.47	0.39	0.19	0.068	0.078	0.035
wa_OptCat_NutMMI	0.47	0.52	0.15	0.066	0.056	0.035
pi_NAWQA_TN_1	0.66	0.67	0.13	0.084	0.028	0.019
pi_Ptpv_TP_all_Hi	0.52	0.71	0.21	0.094	0.032	0.029
x_Shan_e	0.70	0.51	0.25	0.071	0.034	0.027
Median CP diatoms	0.49	0.45	0.18	0.068	0.056	0.035
DO_min	0.63	0.34	0.30	0.066	0.039	0.035
Pmax4hr	0.70	0.37	0.36	0.059	0.099	0.035
Median valid CP BMI, diatoms, & DO	0.50	0.36	0.22	0.067	0.044	0.035
Reference 90th quantile	0.69	0.42	0.30	0.105	0.071	0.054

Change-points for Delta DO and Pmax4hr were calculated based on macroinvertebrate metrics and nutrient concentrations (Table 25). Using nutrient concentrations in the CPA as a response to DO statistics is somewhat circular and might not be an appropriate application of the technique. However, median values for the DO change-points derived from macroinvertebrate metrics were equal to the medians when the nutrient-derived change-points were also included. Change-points were derived for all sites and for the TP Flat sites. In the TP High-Volcanic and TP Steep site classes, there were not enough samples for valid change-point analyses.

Table 25. Change-points (CP) as candidate thresholds from selected benthic macroinvertebrate (BMI) metrics and nutrient concentrations. Values in gray font did not pass the tests for valid change-points.

	Delta DO		Pmax4hr	
Metric	All sites	Flat	All sites	Flat
EPTTax	1.74	1.99	0.358	0.254
EphemTax	1.60	1.88	0.298	0.254
PlecoTax	2.34	2.09	0.275	0.254
IntolTax	2.37	2.42	0.254	0.254
Toler percent	2.46	2.44	0.474	0.439
EPT percent	1.72	2.44	0.290	0.338
Pleco percent	2.02	1.99	0.214	0.214
NonIn percent	2.41	2.42	0.298	0.322
ShredTax	1.56	2.44	0.145	0.214
Clngr percent	2.26	2.44	0.331	0.254
LogTN	5.77	5.73	0.679	0.679
LogTP	2.03	2.06	0.269	0.351
Median	2.30	2.42	0.290	0.254

5.0 Synthesis

The following synthesis of results summarizes the range of candidate thresholds based preceding analyses and workgroup feedback for the purpose of expressing possible thresholds. Thresholds that NMED decides to apply within CWA regulatory programs may deviate from the candidate thresholds presented in this report. NMED should express the basis for threshold decisions in the appropriate place, time, and manner.

5.1 Method strengths and limitations

Three basic methods were used for identifying candidate nutrient thresholds for two nutrients (TN and TP). The methods were applied in three site classes per nutrient and used two biological assemblages (diatoms and macroinvertebrates) as well as DO minimum and maximum productivity (Pmax4hr). The methods included frequency distributions in reference sites, regression interpolation, and change-point analysis.

Candidate thresholds derived from frequency distributions in reference sites are dependent on proper identification of sites with minimal disturbance, classification of those sites to limit natural variability, and selection of an appropriate quantile of the distribution to represent a threshold. The reference sites represent sites with the least disturbance based on reference site criteria and a review processes that allowed quantitative screening of sites based on land uses and activities within the catchment, qualitative evaluation of the sites based on NMED staff familiarity, and identification of individual samples that might have had temporary circumstances with higher than normal disturbance. The site classification based on gradient and location reduced variability in the nutrient conditions. Reference nutrient distributions were quite different among site classes and relatively homogenous within classes.

The quantile selected by NMED was based on preliminary analyses, confidence in reference site selection, reconciliation with existing thresholds, and alignment with stressor-response analyses. The selection of the 90th quantile as opposed to lower quantiles reflects confidence in the reference sites. The 90th quantile of reference values are higher than the current ecoregional thresholds, which range from 0.25 to 0.53 mg/L for TN and 0.02 to 0.09 mg/L for TP (NMED/SWQB 2013; Table 3). In comparison, ranges for the 90th quantiles of reference are from 0.30 to 0.69 mg/L for TN and 0.055 to 0.105 mg/L for TP. The 90th confidence limits around each quantile were shown to further illustrate ranges of possible final nutrient thresholds.

Regression interpolation is a straightforward technique that allowed a direct association of expected response measures to the nutrient conditions. The strength of this approach was its simplicity in direct association through the mean regression equation. The mean regression would best represent nutrient effects if nutrients were the only stressor acting on the biologic assemblages. However, there are many cases where poor biological metrics are associated with low nutrient concentrations, probably due to oligotrophic conditions or unmeasured stressors. These points in the heel of the wedge of a scatter-plot dilute the stressor-response regression, resulting in flatter regression slopes. With a flatter slope, slight changes or inaccuracies in the expected response condition are translated into large differences on the x-axis. Even with a significant regression, interpolated results can be outside of the realistic range of nutrient thresholds.

Change-point analysis was used to determine the point along the x-axis (nutrients) at which the sets of metric values above and below the x value had the least deviance or most precision. The technique can result in change-points even when the change in metrics with nutrients is very gradual. The best use of the technique is with a step-function response, with a clear and consistent change in metric values above and below the change-point. The change-points were evaluated based on indications of a limiting response (quantile regression), precision of the change-point (the breadth of the confidence interval), and coincidence with a local inflection of values (the LOWESS regression line). Even with these evaluation techniques, valid change-points were not always representative of a clear step-function.

Of the two assemblages, a more direct conceptual linkage exists between nutrients and diatoms compared to the indirect linkage between nutrients and macroinvertebrates. It was assumed that diatoms were directly responsive to nutrient conditions. It was also assumed that macroinvertebrates were responsive to secondary effects of nutrients, specifically dissolved oxygen depletion from excess respiration of nutrient-dependent algae, loss of habitat, and change in food quality. However, results of the diatom assemblage were not weighted over results from the macroinvertebrate assemblage. The relationship between nutrients and macroinvertebrates was explored in our analysis of correlations and interactions. There was evidence that nutrients caused changes in DO, but a multiple linkage model (nutrients -> chl-a -> DO -> macroinvertebrates) was less precise than a direct comparison (nutrients -> macroinvertebrates).

Frequency distributions based on the 90th quantile of reference site nutrient values were given the most weight when synthesizing information from all techniques. The stressor-response methods gave valuable supporting or qualifying evidence. Presentation of the 90th quantile of reference sites in the context of all other valid results puts the primary thresholds in the context of thresholds derived through all methods.

5.2 Nutrient and DO Thresholds

Based on the preceding analyses, summaries of candidate nutrient thresholds and ranges of possible final thresholds were compiled (Table 26). The primary thresholds are based on frequency distribution of nutrient values observed in reference sites and site classes. Evidence from the stressor-response analyses general supports the 90th quantile values as thresholds, though medians of valid thresholds developed through change-point analysis and regression interpolation are generally lower than the 90th quantile values. Median stressor-response values are less than the 75th quantile of reference in the TN Flat, TN Moderate, and TP High-

Volcanic classes. In the TN Steep class, the median stressor-response value is near the 80th quantile of reference values. In the TP Flat-Moderate and TP Steep classes, the median stressor-resonse values are greater than or equal to the 90th quantile of reference. The confidence intervals for the reference 90th quantile values describe ranges that are also higher than the stressor-response medians. The CDF curves in the following sections place the reference 90th quantile values within the ranges of stressor-response thresholds.

Table 26. Candidate nutrient threshold values based on frequency distributions and ranges of endpoints by nutrient and site class.

TN		TN Flat	TN Moderate	TN Steep
	Reference 90th quantile	0.69 mg/L	0.42 mg/L	0.30 mg/L
	90% confidence interval	0.62 – 0.85	0.38 – 0.51	0.26 – 0.34
	Stressor-response median	0.52 mg/L	0.33 mg/L	0.26 mg/L
TP		TP High-Volcanic	TP Flat-Moderate	TP Steep
	Reference 90th quantile	0.105 mg/L	0.061 mg/L	0.030 mg/L
	90% confidence interval	0.089 – 0.114	0.051 – 0.069	0.016 – 0.053
	Stressor-response median	0.067 mg/L	0.066 mg/L	0.029 mg/L

For Delta DO values, the reference distribution 90th quantile values were similar in the High-Volcanic and TP Flat site classes at 4 – 5 mg/L DO. Stressor-response analyses were only possible in the TP Flat site class. The regression interpolation using nutrient thresholds were close and slightly lower than the 90th quantile value. Change-point analysis suggested lower thresholds for all macroinvertebrate metrics. Using nutrients in the change-point analysis, only the threshold derived from TN was greater than the reference distribution 90th quantile. Change-point values for all sites were similar to those derived in the TP Flat site class and were generally in the range described by the reference distribution in the TP Steep sites.

Table 27. Threshold ranges for Delta DO derived from reference distributions (Ref Dist 90th), the reference distribution 90% confidence interval (Ref Dist CI90), regression interpolation range (Reg Int range), change-point analysis (CPA) median, and CPA ranges associated with benthic macroinvertebrates (BMI), and nutrients.

	TP High-Volcanic	TP Flat-Moderate	TP Steep	All Classes
Ref Dist 90th	5.02	4.08	1.79	4.16
Ref Dist CI90	3.13 - 7.24	3.52 - 7.27	1.40 - 2.37	3.27-7.13
Reg Int range	NA	3.34 - 4.06	NA	NA
CPA median	NA	2.42	NA	2.30
CPA BMI range	NA	1.88 – 2.44	NA	1.56 – 2.46

CPA nutrient range	NA	2.06 – 5.73	NA	2.03 – 5.77

Results for Pmax4hr thresholds had similar patterns as the Delta DO thresholds, due to the high correlation between the two measures. The reference distribution 90th quantile Pmax4hr values are similar in the High-Volcanic and TP Flat site classes at about 0.65 mg/L DO. Stressor-response analyses were only possible in the TP Flat site class. The regression interpolation using nutrient thresholds were slightly lower than the 90th quantile value. Change-point analysis suggested lower thresholds for all macroinvertebrate metrics. Using nutrients in the change-point analysis, only the threshold derived from TN was greater than the reference distribution 90th quantile. Change-point values for all sites were similar to those derived in the TP Flat site class and were generally in the range described by the reference distribution in the TP Steep sites.

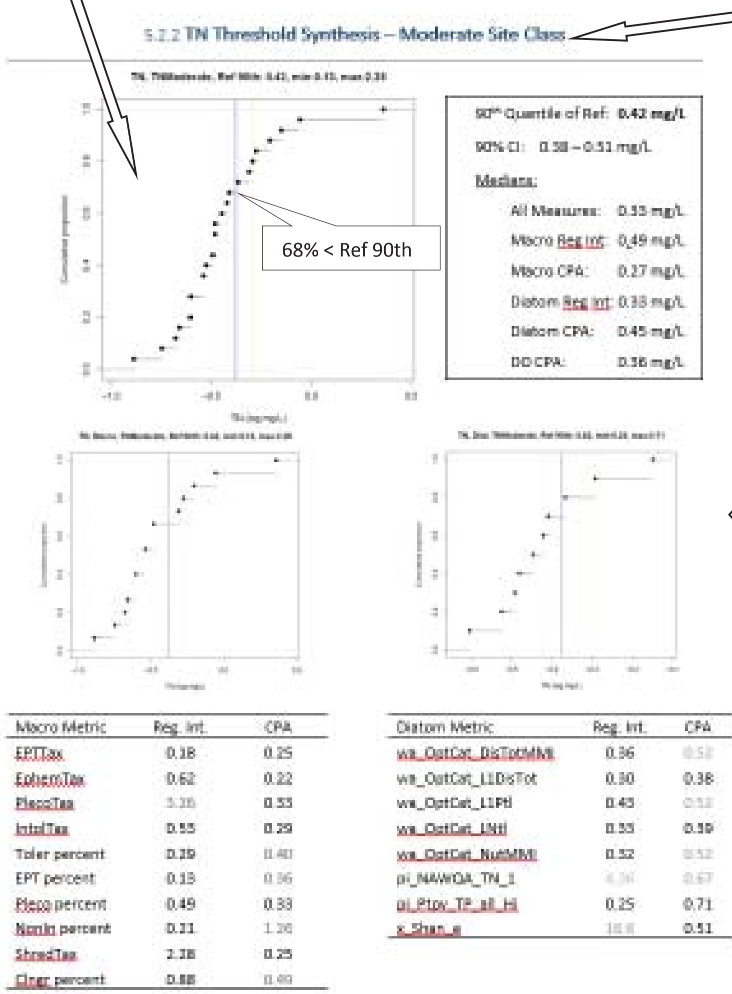
Table 28. Threshold ranges for Pmax4hr derived from reference distributions (Ref Dist 90th), the reference distribution 90% confidence interval (Ref Dist CI90), regression interpolation range (Reg Int range), change-point analysis (CPA) median, and CPA ranges associated with benthic macroinvertebrates (BMI), and nutrients.

	TP High-Volcanic	TP Flat-Moderate	TP Steep	All Classes
Ref Dist 90th	0.635	0.682	0.284	0.659
Ref Dist CI90	0.460 – 0.720	0.493 – 1.200	0.126 – 0.490	0.511-0.688
Reg Int range	NA	0.47 – 0.56	NA	NA
CPA median	NA	0.254	NA	0.290
CPA BMI range	NA	0.214 – 0.439	NA	0.145 – 0.474
CPA nutrient range	NA	0.351 – 0.679	NA	0.269 – 0.679

Candidate Threshold Summary Legend

CDF diagram with all thresholds including 90th quantile of the reference distributions (vertical line with dashed vertical confidence interval) and change-point analysis and regression interpolation from both assemblages. X-axis is log mg/L. Y-axis is proportion of candidate thresholds less than X. The percentage of valid biological thresholds less than the reference 90th quantile is emphasized.

Title, with nutrient and site class



Macro Metric	Reg. Int.	CPA
EPTTax	0.18	0.25
EphemTax	0.62	0.22
PlecoTax	3.26	0.33
IntolTax	0.55	0.29
Toler percent	0.29	0.40
EPT percent	0.13	0.36
Pleco percent	0.49	0.33
Nonin percent	0.21	1.26
ShredTax	2.28	0.25
Clngr percent	0.88	0.49

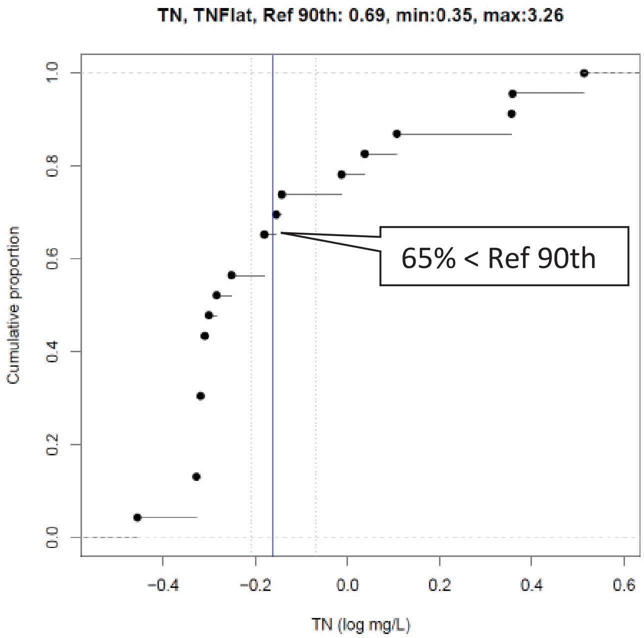
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.36	0.52
wa_OptCat_L1DisTot	0.30	0.38
wa_OptCat_L1Ptl	0.43	0.52
wa_OptCat_LNtl	0.33	0.39
wa_OptCat_NutMMI	0.32	0.52
pi_NAWQA_TN_1	6.36	0.67
pi_Ptov_TP_all_Hi	0.25	0.71
x_Shan_e	10.8	0.51

Summary statistics, with candidate threshold, confidence interval and medians of candidate thresholds by assemblage and method.

CDF diagrams for diatoms and macroinvertebrates, including both methods. Nutrients are shown on a log scale on the axes and are back-transformed in the text.

Threshold values from macroinvertebrate and diatom assemblages and from regression interpolation (Reg. Int.) and change-point analysis (CPA). Gray-shaded values are not valid and were not used in the CDF or summary statistics.

5.2.1 TN Threshold Synthesis – Flat Site Class

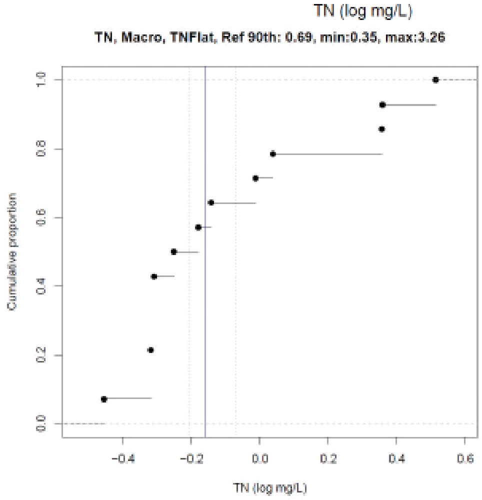


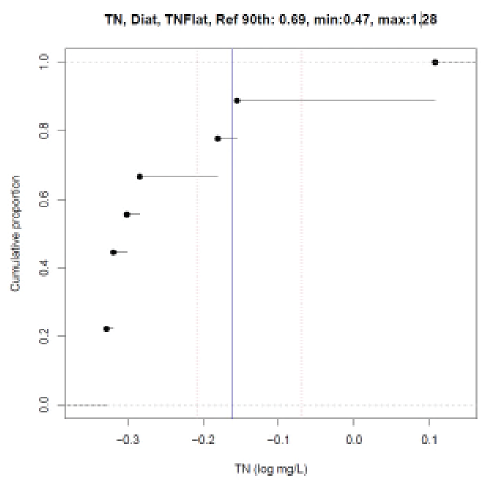
90th Quantile of Ref: **0.69 mg/L**

90% CI: 0.62 – 0.85 mg/L

<u>Medians:</u>

All Measures:	0.52 mg/L
Macro Reg Int:	2.28 mg/L
Macro CPA:	0.53 mg/L
Diatom Reg Int:	1.28 mg/L
Diatom CPA:	0.49 mg/L
DO CPA:	NA

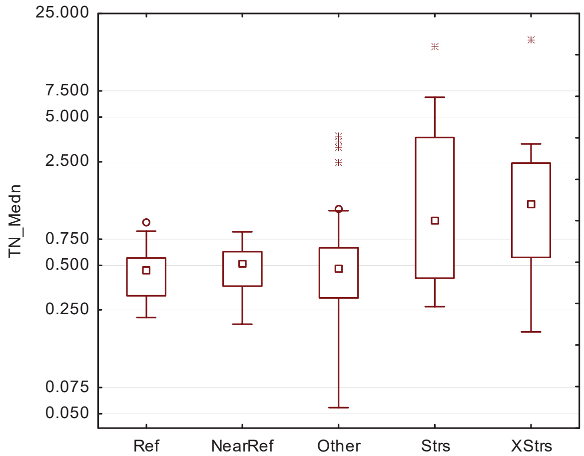




Macro Metric	Reg.Int.	CPA
EPTTax	3.70	0.49
EphemTax	2.27	0.49
PlecoTax	3.26	0.56
IntolTax	3.48	0.48
Toler percent	500.8	0.66
EPT percent	398.7	0.97
Pleco percent	0.49	0.35
NonIn percent	28.3	0.72
ShredTax	2.28	0.48
Clngr percent	108.6	1.09

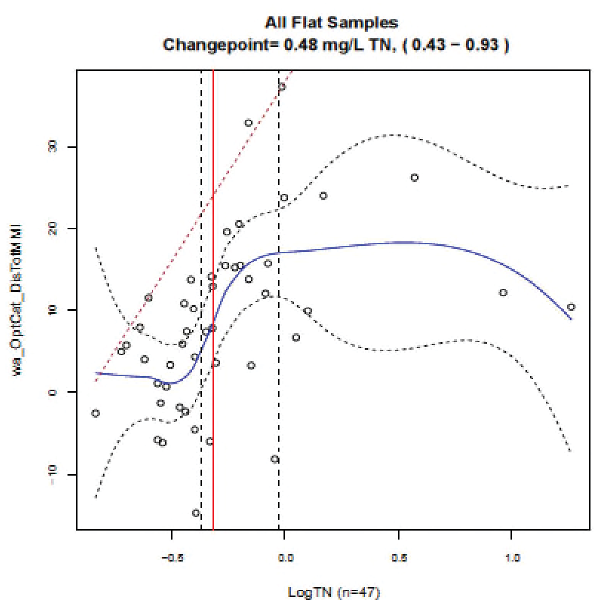
Diatom Metric	Reg.Int.	CPA
wa_OptCat_DisTotMMI	10.4	0.48
wa_OptCat_L1DisTot	18.3	0.50
wa_OptCat_L1Ptl	7.45	0.48
wa_OptCat_LNtl	10.5	0.47
wa_OptCat_NutMMI	9.26	0.47
pi_NAWQA_TN_1	1.28	0.66
pi_Ptpv_TP_all_Hi	7.98	0.52
x_Shan_e	2.26	0.70

TN Threshold Synthesis – Flat Site Class (continued)



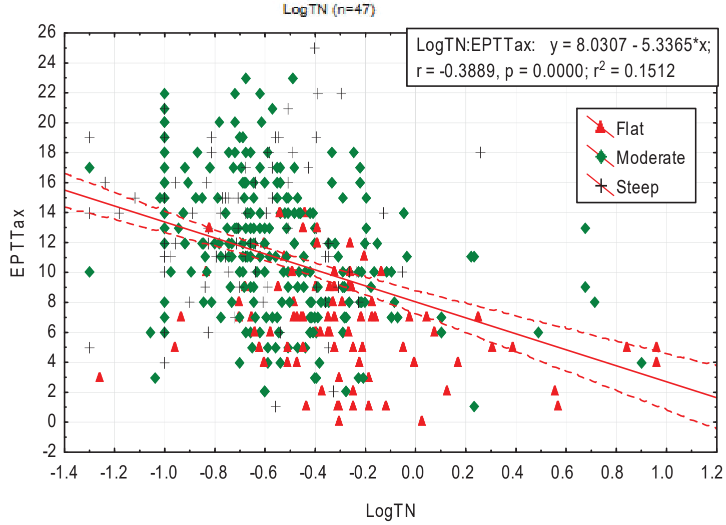
TN values were lowest in reference sites in the TN Flat site class and increased most in the stressed and extremely stressed sites (Figure 45). More than half of the stressed and extremely stressed sites were greater than the 90th quantile value (0.69 mg/L TN). Also see Section 4.2.

Figure 45. Site median TN value distributions along the disturbance gradient for sites in the TN Flat site class.



Several change-points were identified near 0.50 mg/L for both diatoms and macroinvertebrate metrics (e.g., Figure 46). Also see Section 4.5 and Appendix L.

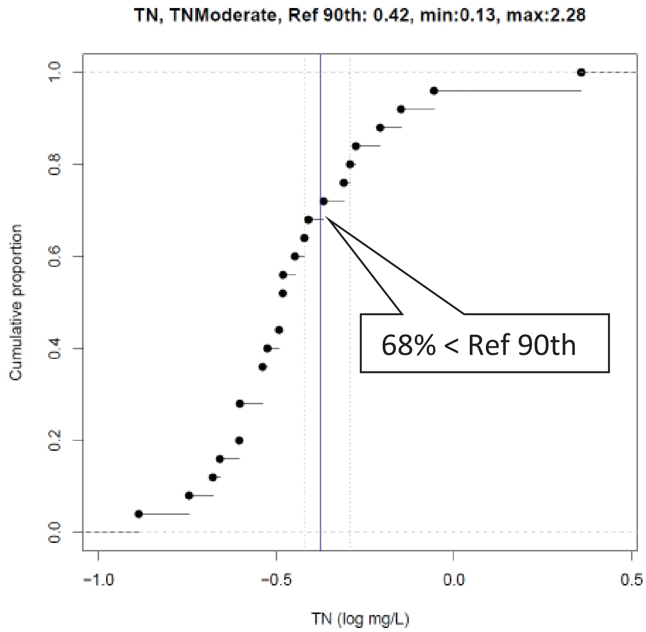
Figure 46. Change-point plot for TN and the weighted average disturbance index diatom metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



While all but one metric had a significant overall regression equation (e.g., Figure 47), the reference quartiles of the metrics in the TN Flat sites were substantially different than those in the other site classes. This resulted in high interpolated values for TN. Also see Section 4.4 and Appendix K.

Figure 47. Regression plot for TN and EPT taxa. In the TN Flat site class, the reference quartile for EPT taxa was 5 taxa, which translates to 3.7 mg/L TN.

5.2.2 TN Threshold Synthesis – Moderate Site Class

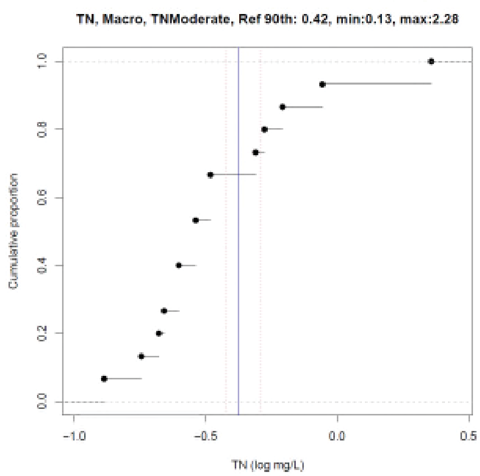


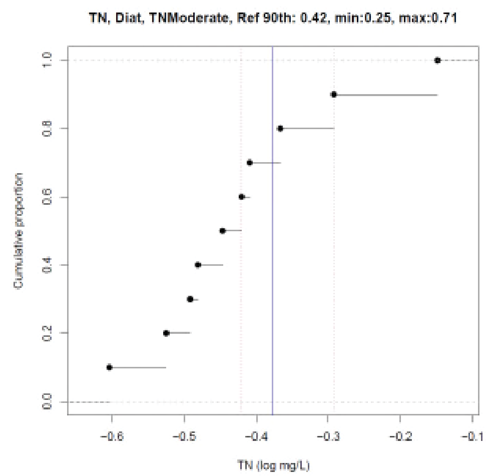
90th Quantile of Ref: **0.42 mg/L**

90% CI: 0.38 – 0.51 mg/L

Medians:

All Measures:	0.33 mg/L
Macro Reg Int:	0.49 mg/L
Macro CPA:	0.27 mg/L
Diatom Reg Int:	0.33 mg/L
Diatom CPA:	0.45 mg/L
DO CPA:	0.36 mg/L

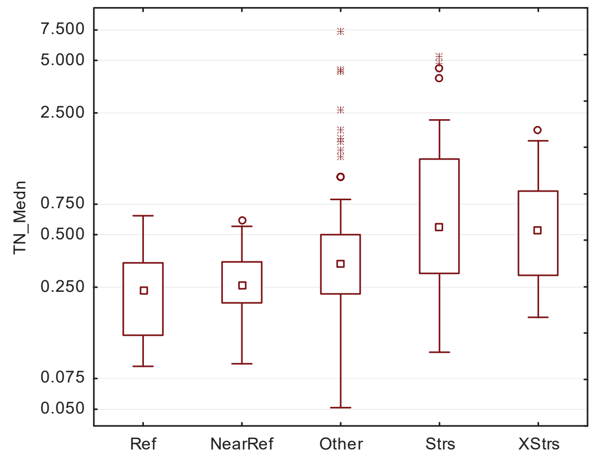




Macro Metric	Reg. Int.	CPA
EPTTax	0.18	0.25
EphemTax	0.62	0.22
PlecoTax	3.26	0.33
IntolTax	0.53	0.29
Toler percent	0.29	0.40
EPT percent	0.13	0.36
Pleco percent	0.49	0.33
NonIn percent	0.21	1.26
ShredTax	2.28	0.25
Clngr percent	0.88	0.49

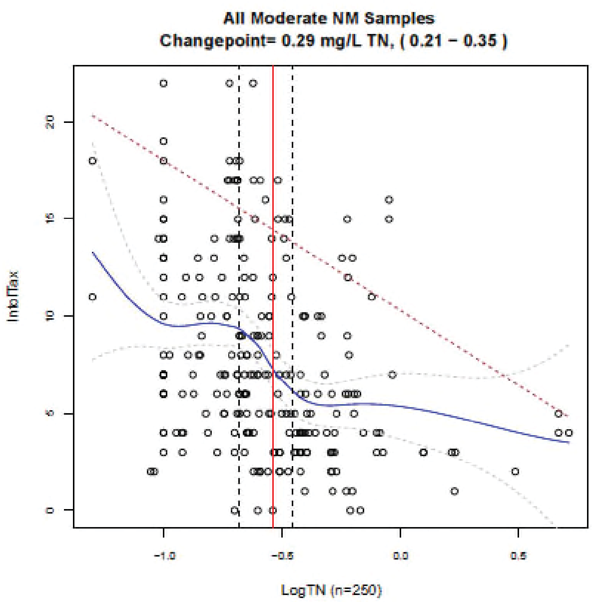
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.36	0.52
wa_OptCat_L1DisTot	0.30	0.38
wa_OptCat_L1Ptl	0.43	0.52
wa_OptCat_LNtl	0.33	0.39
wa_OptCat_NutMMI	0.32	0.52
pi_NAWQA_TN_1	4.36	0.67
pi_Ptpv_TP_all_Hi	0.25	0.71
x_Shan_e	16.6	0.51

TN Threshold Synthesis – Moderate Site Class (continued)



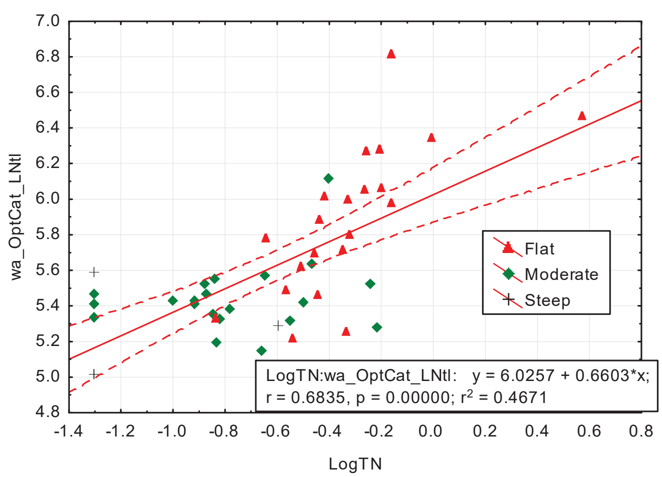
TN values were lowest in reference sites in the TN Moderate site class and increased most in the stressed and extremely stressed sites (Figure 48). More than half of the stressed and extremely stressed sites were greater than the 90th quantile value (0.42 mg/L TN). Also see Section 4.2.

Figure 48. Site median TN value distributions along the disturbance gradient for sites in the TN Moderate site class.



Most change-points were identified at TN values slightly less than the 90th quantile of reference sites (e.g., Figure 49). Also see Section 4.5 and Appendix L.

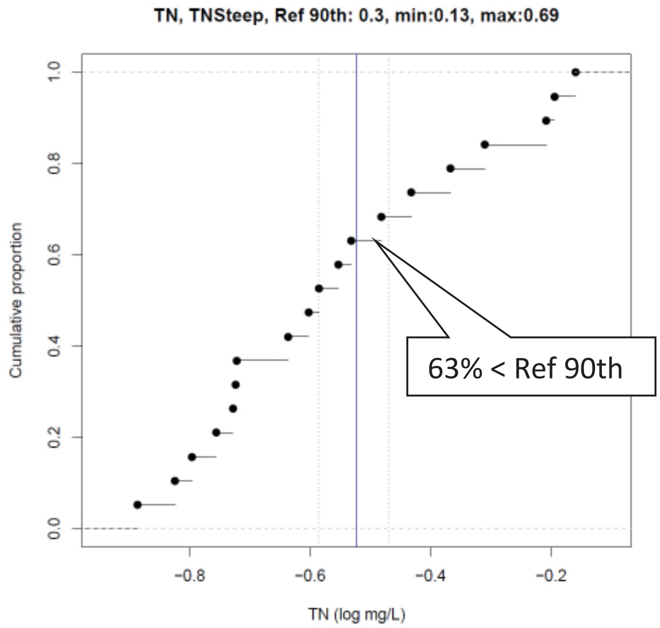
Figure 49. Change-point plot for TN and the intolerant taxa macroinvertebrate metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



All but one metric had a significant overall regression equation (e.g., Figure 50). The reference quartiles of metrics in the TN Moderate sites were similar to those in the Steep site class. Also see Section 4.4 and Appendix K.

Figure 50. Regression plot for TN and weighted average diatom nitrogen sensitivity. In the TN Moderate site class, the reference quartile for the metric was 6.0, which translates to 0.33 mg/L TN.

5.2.3 TN Threshold Synthesis – Steep Site Class

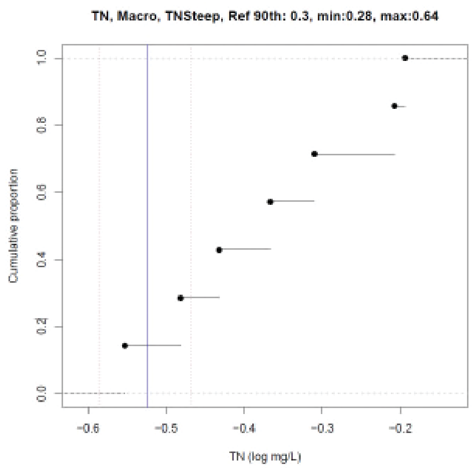


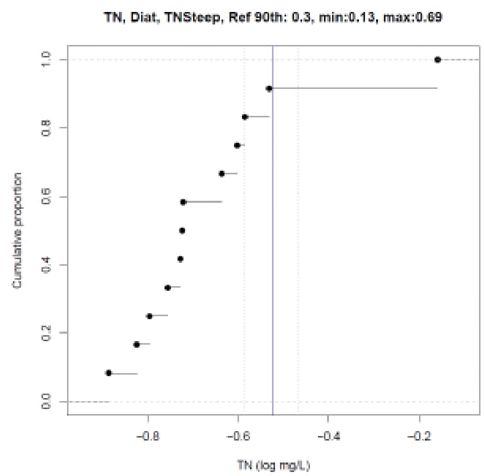
90th Quantile of Ref: **0.30 mg/L**

90% CI: 0.26 – 0.34 mg/L

Medians:

All Measures:	0.26 mg/L
Macro Reg Int:	0.46 mg/L
Macro CPA:	0.21 mg/L
Diatom Reg Int:	0.23 mg/L
Diatom CPA:	0.18 mg/L
DO CPA:	0.30 mg/L

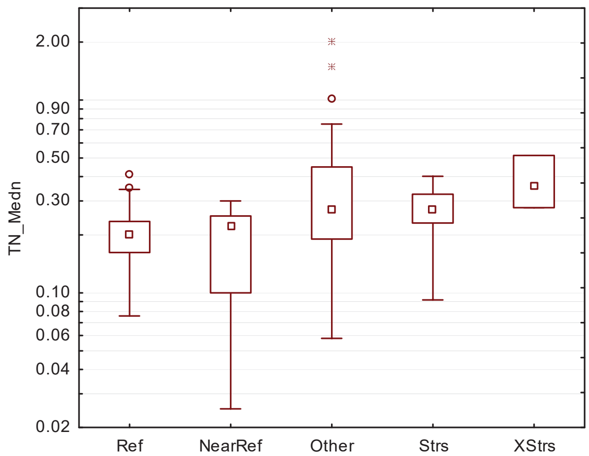




Macro Metric	Reg. Int.	CPA
EPTTax	0.43	0.42
EphemTax	0.62	0.28
PlecoTax	3.26	0.25
IntolTax	0.85	0.39
Toler percent	0.33	0.26
EPT percent	21.5	0.22
Pleco percent	0.49	0.14
NonIn percent	0.37	0.23
ShredTax	0.64	0.23
Clngr percent	2.44	0.28

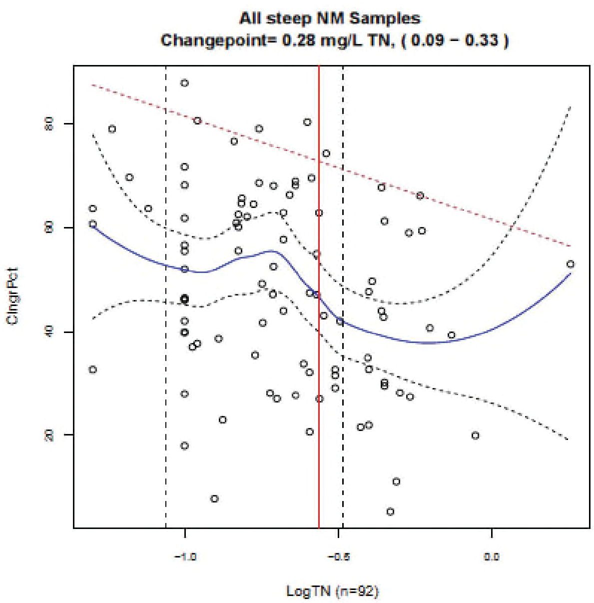
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.19	0.16
wa_OptCat_L1DisTot	0.19	0.26
wa_OptCat_L1Ptl	0.29	0.13
wa_OptCat_LNtl	0.18	0.19
wa_OptCat_NutMMI	0.23	0.15
pi_NAWQA_TN_1	5.32	0.13
pi_Ptpv_TP_all_Hi	0.69	0.21
x_Shan_e	>10,000	0.25

TN Threshold Synthesis – Steep Site Class (continued)



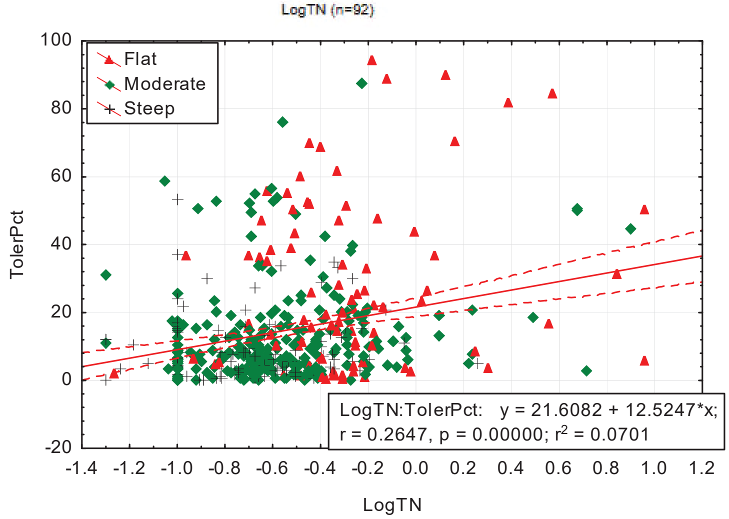
TN values were lowest in reference sites in the TN Steep site class and increased gradually with increasing stress (Figure 51). Approximately half of the stressed and extremely stressed sites were greater than the 90th quantile value (0.30 mg/L TN). Also see Section 4.2.

Figure 51. Site median TN value distributions along the disturbance gradient for sites in the TN Steep site class.



Most change-points were identified at TN values slightly less than the 90th quantile of reference sites. Only one macroinvertebrate metric gave acceptable change-point results (Figure 52). Also see Section 4.5 and Appendix L.

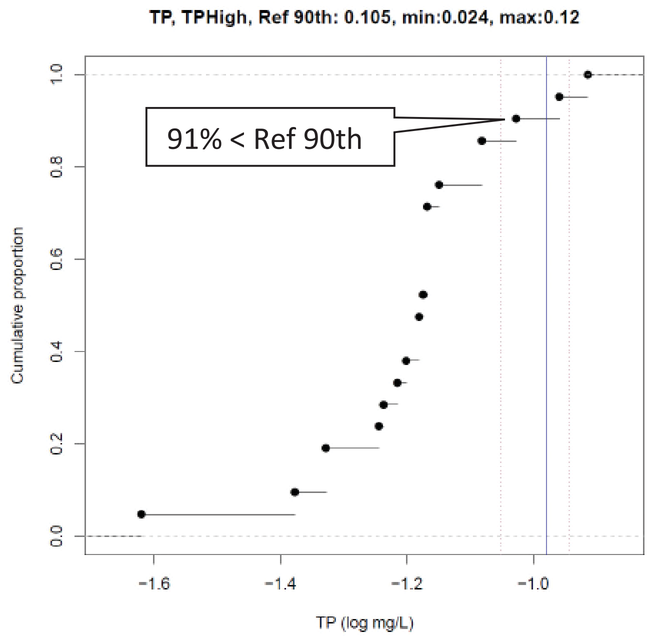
Figure 52. Change-point plot for TN and the percent clinger macroinvertebrate metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



All but one metric had a significant overall regression equation (e.g., Figure 53). Interpolation results closest to the reference quantile value were for the macroinvertebrate percent tolerance metric. Also see Section 4.4 and Appendix K.

Figure 53. Regression plot for TN and macroinvertebrate percent tolerance. In the TN Steep site class, the reference quartile for the metric was 15.5%, which translates to 0.33 mg/L TN.

5.2.4 TP Threshold Synthesis – High-Volcanic Site Class

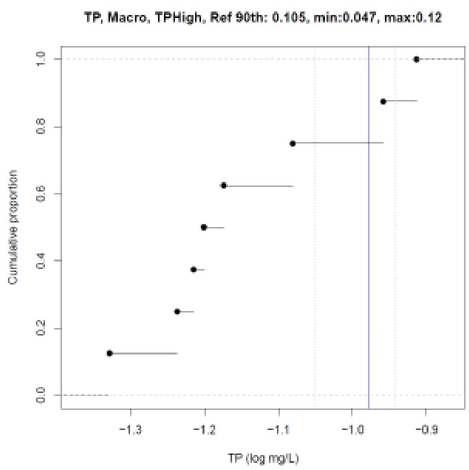


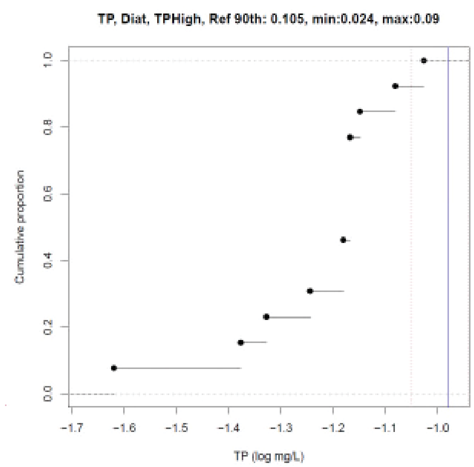
90[th] Quantile of Ref: **0.105 mg/L**

90% CI: 0.89 – 0.114 mg/L

Medians:

All Measures:	0.067 mg/L
Macro Reg Int:	0.110 mg/L
Macro CPA:	0.063 mg/L
Diatom Reg Int:	0.052 mg/L
Diatom CPA:	0.068 mg/L
DO CPA:	0.059 mg/L

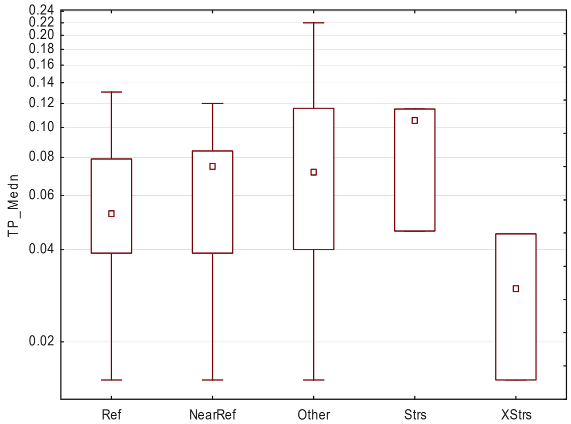




Macro Metric	Reg. Int.	CPA
EPTTax	0.11	0.067
EphemTax	1.00	0.058
PlecoTax	1.61	0.063
IntolTax	0.88	0.061
Toler percent	0.22	0.083
EPT percent	59,102	0.047
Pleco percent	1.74	0.114
NonIn percent	1.46	0.083
ShredTax	56.47	0.047
Clngr percent	13.22	0.122

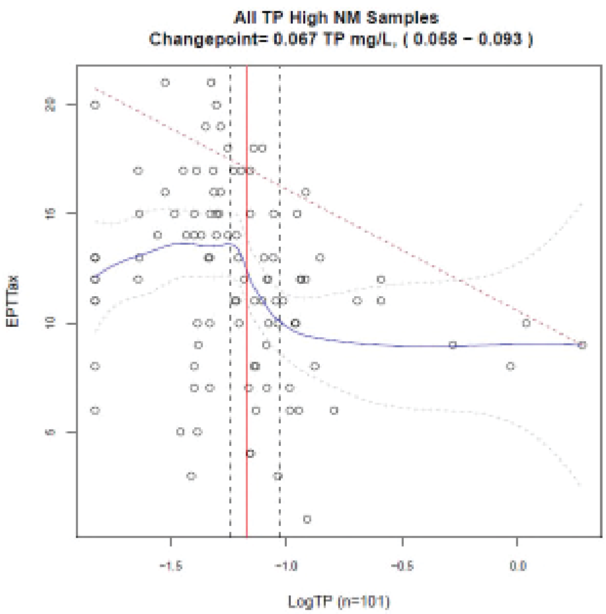
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.042	0.068
wa_OptCat_L1DisTot	0.024	0.068
wa_OptCat_L1Ptl	0.068	0.066
wa_OptCat_LNtl	0.057	0.068
wa_OptCat_NutMMI	0.047	0.066
pi_NAWQA_TN_1	0.457	0.084
pi_Ptpv_TP_all_Hi	0.083	0.094
x_Shan_e	9.272	0.071

TP Threshold Synthesis – High-Volcanic Site Class (continued)



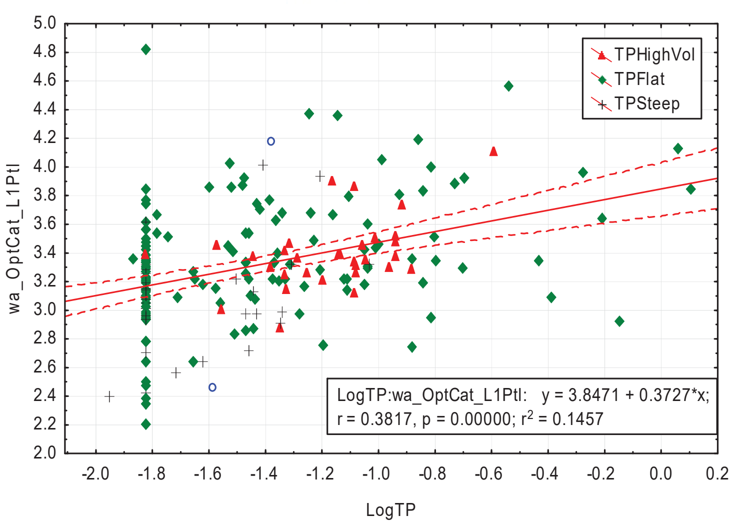
Median TP values in reference sites of the TP High-Volcanic site class were lower than those in the stressed sites and higher than those in the few extremely stressed sites (Figure 54). Most of the sites that were greater than the 90th quantile value (1.05 mg/L TP) were in the Other category. Also see Section 4.2.

Figure 54. Site median TP value distributions along the disturbance gradient for sites in the TP High-Volcanic site class.



Most change-points were identified at TP values less than the 90th quantile of reference sites, like the EPT taxa response (Figure 55). Only the percent clinger macroinvertebrate metric had a higher change-point. Also see Section 4.5 and Appendix L.

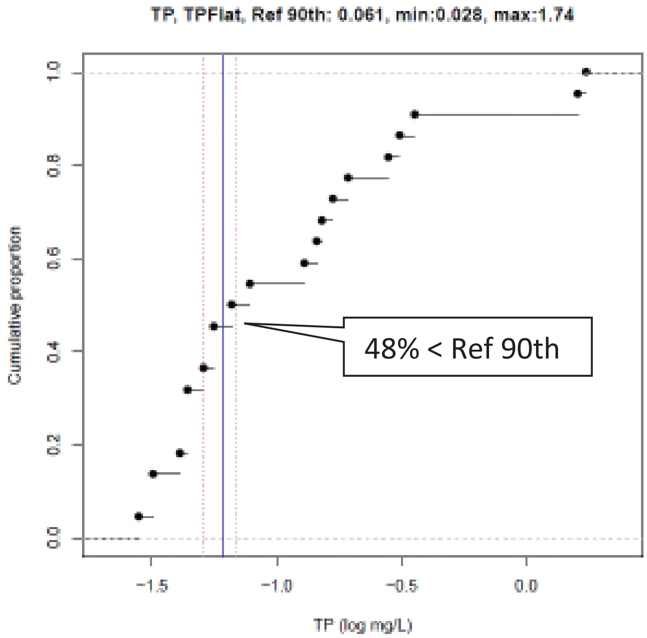
Figure 55. Change-point plot for TP and the EPT taxa macroinvertebrate metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



All but one diatom metric had interpolated TP values less than the reference 90th quantile, such as the weighted average phosphorus diatom metric (Figure 56). Also see Section 4.4 and Appendix K.

Figure 56. Regression plot for TP and weighted average diatom phosphorus sensitivity. In the TP High Volcanic sites, the metric upper reference quartile was 3.4, which translates to 0.068 mg/L TP.

5.2.5 TP Threshold Synthesis – Flat-Moderate Site Class

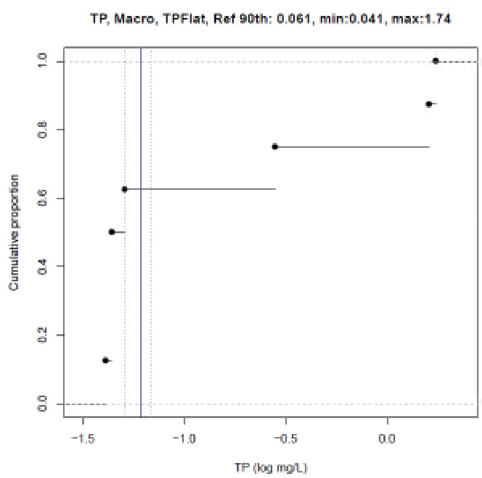


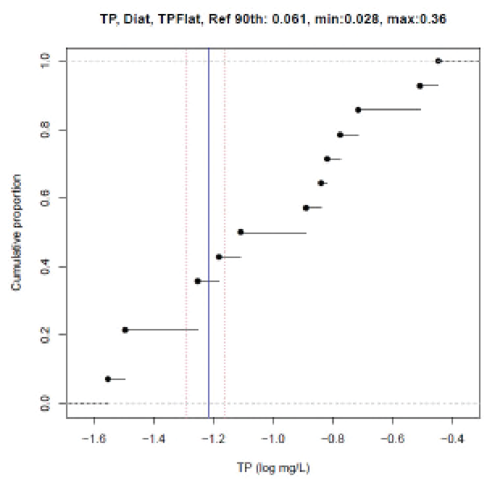
90th Quantile of Ref: **0.061 mg/L**

90% CI: 0.051 – 0.069 mg/L

Medians:

All Measures:	0.066 mg/L
Macro Reg Int:	1.675 mg/L
Macro CPA:	0.044 mg/L
Diatom Reg Int:	0.168 mg/L
Diatom CPA:	0.056 mg/L
DO CPA:	0.099 mg/L

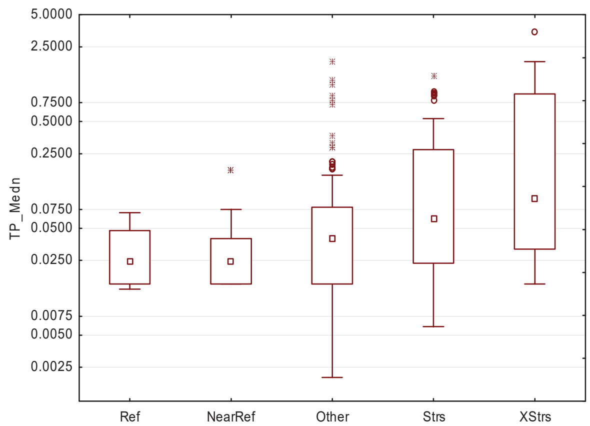




Macro Metric	Reg. Int.	CPA
EPTTax	3.39	0.044
EphemTax	211	0.044
PlecoTax	1.61	0.041
IntolTax	6.22	0.051
Toler percent	3.11	0.052
EPT percent	31.91	0.014
Pleco percent	1.74	0.044
NonIn percent	0.281	0.014
ShredTax	56.47	0.151
Clngr percent	5.01	0.051

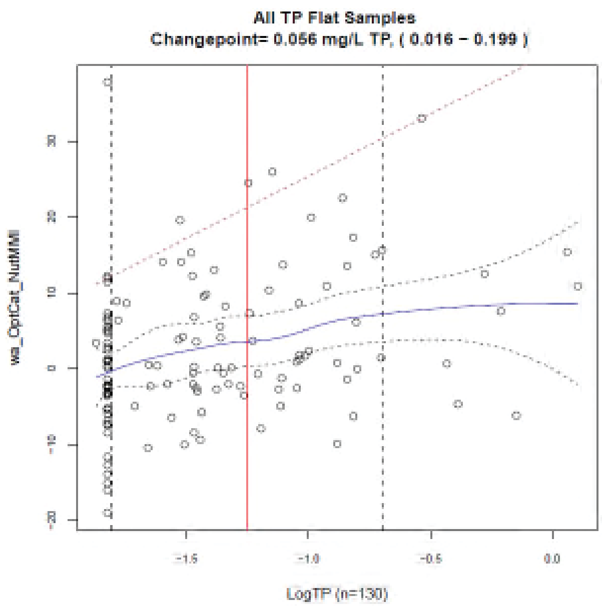
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.168	0.056
wa_OptCat_L1DisTot	0.358	0.066
wa_OptCat_L1Ptl	0.145	0.032
wa_OptCat_LNtl	0.311	0.078
wa_OptCat_NutMMI	0.193	0.056
pi_NAWQA_TN_1	0.129	0.028
pi_Ptpv_TP_all_Hi	0.152	0.032
x_Shan_e	7.272	0.034

TP Threshold Synthesis – Flat-Moderate Site Class (continued)



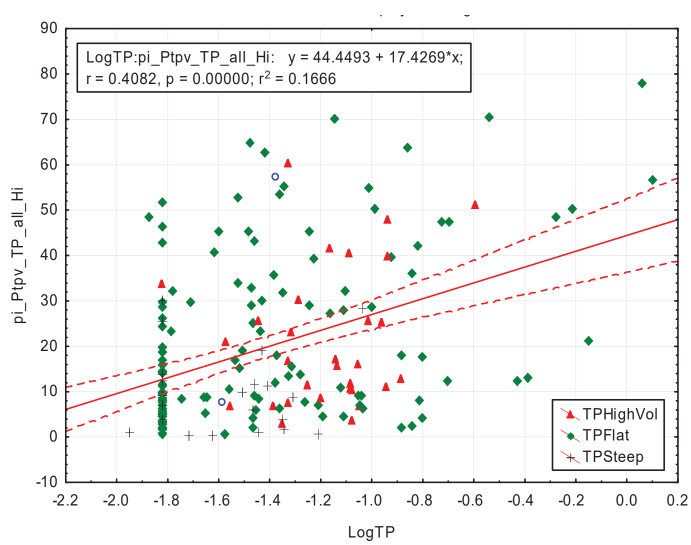
TP values were lowest in reference sites in the TP Flat-Moderate site class and increased gradually with increasing stress (Figure 57). More than half of the stressed and extremely stressed sites were greater than the 90th quantile (0.071 mg/L TP). Also see Section 4.2.

Figure 57. Site median TN value distributions along the disturbance gradient for sites in the TP Flat-Moderate site class.



Change-points for macroinvertebrates were lower than the 90th quantile of reference sites. For diatoms, valid change-points bracketed the 90th quantile. The weighted average nutrient index diatom metric had a change-point close to the reference 90th quantile (Figure 58). Also see Section 4.5 and Appendix L.

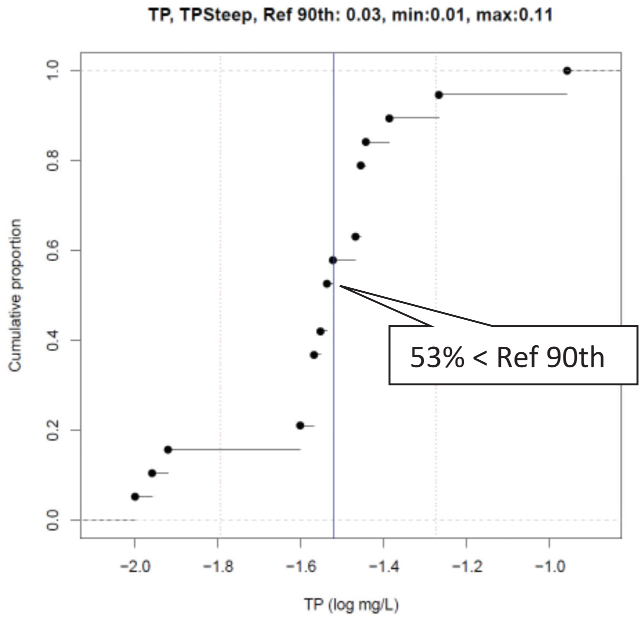
Figure 58. Change-point plot for TP and the weighted average nutrient index diatom metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



Most regression interpolations for macroinvertebrate metrics were unreasonably high. Diatom results were also higher than the 90th quantile of reference sites (e.g., Figure 59). Also see Section 4.4 and Appendix K.

Figure 59. Regression plot for TP and percent TP tolerant diatom metric. In the TP Flat-Moderate site class, the upper reference quartile for the metric was 30.2, which translates to 0.152 mg/L TP.

5.2.6 TP Threshold Synthesis – Steep Site Class

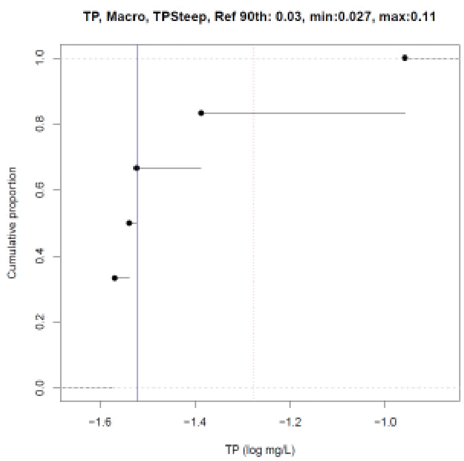


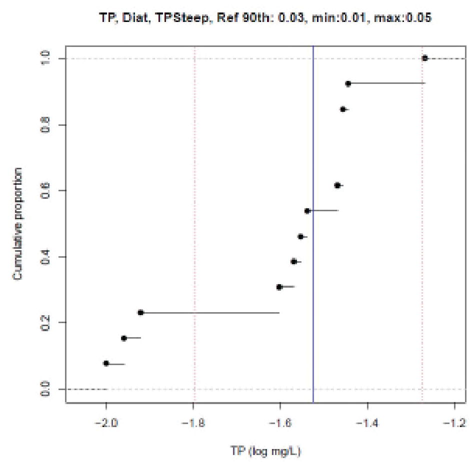
90[th] Quantile of Ref: **0.030 mg/L**

90% CI: 0.016 – 0.053 mg/L

<u>Medians:</u>

All Measures:	0.029 mg/L
Macro Reg Int:	0.110 mg/L
Macro CPA:	0.029 mg/L
Diatom Reg Int:	0.026 mg/L
Diatom CPA:	0.035 mg/L
DO CPA:	0.035 mg/L

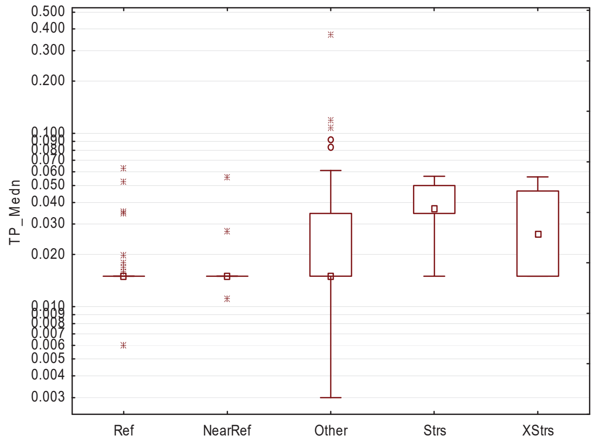




Macro Metric	Reg. Int.	CPA
EPTTax	0.11	0.030
EphemTax	1.00	0.030
PlecoTax	0.15	0.027
IntolTax	0.33	0.029
Toler percent	0.017	0.041
EPT percent	491758	0.029
Pleco percent	0.80	0.027
NonIn percent	0.003	0.018
ShredTax	0.60	0.017
Clngr percent	0.50	0.022

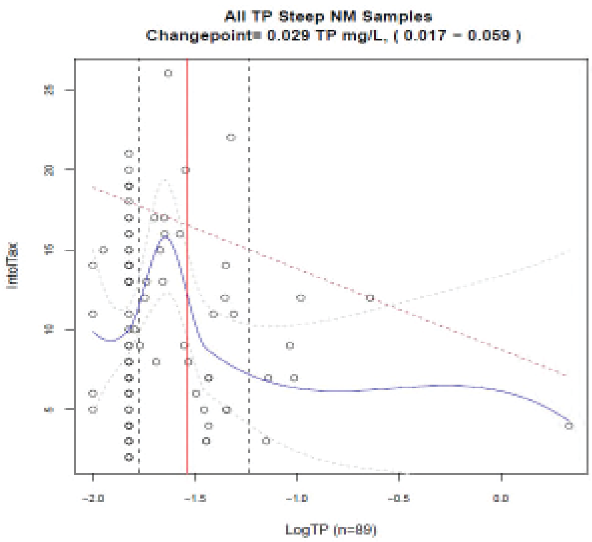
Diatom Metric	Reg. Int.	CPA
wa_OptCat_DisTotMMI	0.028	0.035
wa_OptCat_L1DisTot	0.027	0.034
wa_OptCat_L1Ptl	0.029	0.036
wa_OptCat_LNtl	0.054	0.035
wa_OptCat_NutMMI	0.025	0.035
pi_NAWQA_TN_1	0.010	0.019
pi_Ptpv_TP_all_Hi	0.011	0.029
x_Shan_e	0.012	0.027

TP Threshold Synthesis – Steep Site Class (continued)



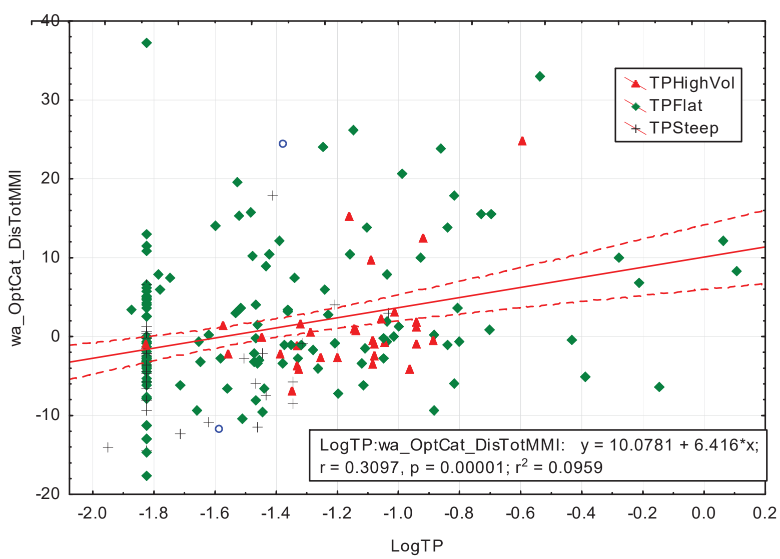
TP values were lowest in reference sites in the TP Steep site class and median values were highest in the stressed and extremely stressed sites (Figure 60). The 90th quantile value (0.054 mg/L TP) is exceeded most often in sites in the Other category. Also see Section 4.2.

Figure 60. Site median TP value distributions along the disturbance gradient for sites in the TP Steep site class.



Relatively few change-points were valid, but those that were fell near the 90th quantile of reference sites (e.g., Figure 61). Also see Section 4.5 and Appendix L.

Figure 61. Change-point plot for TP and the intolerant taxa macroinvertebrate metric, showing the change-point with confidence intervals (vertical solid and dashed lines), the 95th quantile regression line (slanting dashed line), and the LOWESS regression (blue fitted line).



Most regression interpolations for macroinvertebrate metrics were high. Diatom results were closer to the 90th quantile (e.g., Figure 62). Also see Section 4.4 and Appendix K.

Figure 62. Regression plot for TP and weighted average disturbance index metric. In the TP Steep site class, the reference quartile for the metric was 1.1, which translates to 0.028 mg/L TP.

6.0 Discussion

The goals of this report were to describe the intent, methods, and results of nutrient threshold development for perennial wadeable streams in New Mexico. The basis of the threshold development process was the Empirical Approaches for Nutrient Criteria Derivation (USEPA 2009). This approach includes five steps, (1) Selecting and Evaluating Data; (2) Assessing the Strength of the Cause-Effect Relationship; (3) Analyzing Data; (4) Evaluating Estimated Stressor-Response relationship; and (5) Evaluating Candidate Stressor-Response Criteria. Quantitative nutrient translator values that were derived using reference conditions and stressor-response analyses will be used to apply the existing narrative nutrient criterion. The narrative criterion is to protect against plant nutrient levels associated with “undesirable aquatic life” and “dominance of nuisance species”. Through the conceptual model and analyses of select linkages within it, this report suggested ranges of numeric thresholds for TP and TN that could be associated with reference sites and protection of aquatic life. Using the nutrient threshold ranges and other evidence, NMED SWQB can select and incorporate revised numeric thresholds for TN and TP into their assessment protocols and refine thresholds for other water quality variables.

Nutrient thresholds ranges were derived within site classes with relatively homogenous expectations for natural nutrient conditions. The least disturbed reference sites throughout the state had variable nutrient concentrations. The variation was associated with land slope in the catchment of each site. For TP geologic regions as well as land slope was used to identify site classes and areas with high natural phosphorus that can be associated with volcanic geology. Three nutrient-specific site classes were identified for TP and TN. While residual variability in nutrient concentrations was evident after site classification, the site classes sufficiently reduced the overall statewide variability for threshold derivation. The site classes and ranges of associated nutrient thresholds are summarized in Tables 11 and 26.

The TN and TP site classes are not directly comparable to the ecoregional classes used in previous nutrient threshold development efforts in NM. However, the thresholds derived for the new site classes using distribution statistics and stressor response analysis (Table 26) are generally higher than those currently used (Table 3). For TN, the range of thresholds derived in this report is from 0.26 to 0.69 mg/L. This compares to a range of 0.25 to 0.53 mg/L TN currently used. For TP, the derived thresholds are from 0.029 to 0.105 mg/L compared to the currently used thresholds of 0.020 to 0.090 mg/L TP.

The synthesis of threshold ranges (Section 5.0) compared the reference frequency distribution values and confidence limits with the ranges and central tendencies of the stressor-response values. NMED SWQB can select nutrient thresholds that are both representative of most reference sites in a site class and that are protective of meaningful biological conditions. The selection should weigh the merits of each of the analytical methods for threshold derivation, the endpoints (reference conditions, aquatic assemblage, or DO measure), and the ranges of valid thresholds.

The thresholds derived from the 90th quantile of reference nutrient concentrations in each site class are primary thresholds under consideration. The emphasis is based on high confidence in the reference site designations. Reliance on the 90th quantile of the reference distribution is justified if the reference sites are correctly identified. NMED is relatively certain that the reference sites are identified correctly because of the rigorous technical analysis of disturbance variables and the qualitative review process. With less confidence, a lower quantile of the distribution (e.g., 75th) would be selected. The 75th quantile values for TN and TP in each site class were consistently near but lower than the lower confidence intervals for the 90th quantile values (Table 12). NMED should select appropriate quantiles to use for each site class based on comparison of the confidence intervals and range of candidate thresholds as summarized in the cumulative frequency distributions, as well as the merits of each of the analytical methods.

The medians of the valid stressor-response nutrient thresholds were within the 90% confidence interval of the reference-derived 90th quantile thresholds in the Steep site classes for both TN and TP and in the Flat-Moderate site class for TP (Table 26). In the TN Flat, TN Moderate, and the TP High-Volcanic classes, the medians of the stressor-response thresholds were lower than the lower confidence limit for the reference derived thresholds. In the TN Flat and TN Moderate classes, about 70% of the valid stressor-response thresholds were less than the reference derived thresholds (See CDFs in Section 5.2). In the TP High-Volcanic site class, about 90% of the stressor-response thresholds for both macroinvertebrates and diatoms were less than the reference derived threshold. The aquatic assemblages are apparently responsive to increases in nutrients even when those increases might not be associated with identifiable human disturbance in the landscape. In this case, NMED might weigh the feasibility of reducing nutrients to protect biotic assemblage metrics (using a lower threshold) in an environment with ample nutrients as a natural condition (suggested by the higher reference-derived threshold).

Limitations of analytical methods

Regressions are used to estimate a relationship between any pair of variables. When used for threshold derivation, it is important to consider the theoretical assumptions underlying regression inferences. More specifically, one must assess: (1) whether the assumed linear functional form is sufficiently representative of the actual relationship, (2) whether the sampling variability in the dependent variable is distributed as assumed, (3) whether the magnitude of the sampling variability in the dependent variable changes across the range of predictions, and (4) whether the samples used to fit the model are independent of one another. In the regression interpolation analysis, these assumptions were based on regression statistics and examination of the scatter plots. To address representativeness, the regression was required to have significant p-value (<0.05). Flat slopes (low r^2) did not automatically disqualify a regression relationship from being used for derivation of thresholds. Variability around the regression line was not formally assess, but scatter plots were reviewed and typically showed a wedge-shaped plot with higher variability in response variables at lower stressor values compared to higher stressor values. This is assumed to be related to unmeasured stressors that do not appear in the x-axis, but still limit the response. The samples were assumed to be independent of each other because they were from unique sites and times.

The regression interpolation analyses resulted in valid thresholds for TN in the Moderate and Steep site classes. In the TN Flat class, regression interpolation results were high or outside of the acceptable range of TP values for both macroinvertebrates and diatoms. For TP, regression interpolation resulted in valid thresholds when derived from diatom metrics, but not for most macroinvertebrate metrics. The unacceptably high nutrient thresholds derived from the regression interpolation method may be attributed to several factors, including the wedge shape of the stressor-response relationships, the high variability observed in the modifying factors (Section 4.3), and the uncertainty in defining critical response levels based on reference quartiles. Potential thresholds were considered valid if they were within the general observed range of values in each site class. The high ends of potential threshold ranges were usually attributed to a result from the regression interpolation method.

The change-point analysis identifies nutrient concentrations that are associated with high degrees of change in the response variable. The change-points typically fall in the middle of the steepest parts of the LOWESS regression curves. Nutrient stressor-response effects might be occurring at the first inflection of the curve instead of at its steepest point. Therefore, the change-point might be more indicative of more severe effects instead of the earliest minimal effects. The change-point analysis was mostly invalid for deriving TN thresholds based on macroinvertebrates in the TN Steep class. In other classes and for all classes using diatoms, valid thresholds were derived from at least half of the metrics.

Conceptual relationships supported by the analyses

TP and benthic chl-a were positively correlated in all sites and in the TP High-Volcanic site class, supporting the conceptual linkage between higher nutrients and higher chl-a levels. TN and benthic chl-a were not significantly correlated in any of the data subsets, suggesting that TP is the limiting nutrient in most cases. Nutrient and chl-a correlations were never significant in the Steep site classes, in which nutrient concentrations were relatively low. Besides having a short stressor gradient, steep streams might also have scouring flows and closed canopies that are not optimal conditions for chl-a growth. Benthic chl-a was also related to conductivity, elevation, drainage area, latitude, longitude, and pH, though these were not successfully used to reduce variability and increase the correlations between nutreints and chl-a.

Sestonic chl-a was significantly related to both TN and TP in the NRSA sites, except in the steep sites. In slow, broad streams and small rivers, the chl-a in the water column might actually be planktonic. However, it is also likely that much of the sestonic chl-a is of benthic origin, having sluffed off due to overgrowth, grazing, or scouring. For sestonic chl-a, correlations with TN were stronger than those with TP. Sestonic chl-a was also related to elevation, conductivity, turbidity, and drainage area.

The conceptual model included a relationship between chl-a and DO dynamics, in which increased chl-a was related to lower minimum DO and greater DO flux over the course of a day.

While benthic chl-a was correlated with changes in DO (DeltaDO, Pmax4hr, Rmax4hr, and ER), it was not correlated to minimum DO. This might be due to production and respiration of the living algal biomass, whereas the concept was that excessive chl-a would also result in high respiration and low DO caused by dead biomass and high organic content in the sediments. The lack of low minimum DO in relation to high chl-a and the association of high sestonic chl-a with high nutrients suggests that algae that overgrows and sluffs does not end up in the sediments, but is transported downstream. High nutrients were related to low minimum DO, confirming an indirect relationship that does not depend on chl-a.

The diatom assemblage was expected to change in response to increases in nutrients. As expected, several diatom metrics were correlated with TN and TP. TN was positively correlated with conductivity, which was also commonly correlated with metric values. Based on numbers of significant correlations, diatoms appear to be more sensitive to TP than to TN. When other variables were allowed in multiple regression models, conductivity often became the dominant predictor. Conductivity (nor other modifying factors) was not factored out in the analysis of nutrient-diatom relationships. Eight diatom metrics were selected to characterize responses to nutrients.

The expectation in the conceptual model was that benthic macroinvertebrates would respond to increased nutrients because of the effects of increase algal biomass (chl-a) and decreased DO. There were 14-16 macroinvertebrate metrics that were significantly correlated with chl-a and diel DO measures, including minimum DO and Pmax4hr. However, though the relationship between nutrients and macroinvertebrates is conceptually indirect, more macroinvertebrate metrics were significantly correlated to TN and TP than to chl-a or DO measures. Ten macroinvertebrate metrics were selected to characterize responses to nutrients.

Dissolved Oxygen

DO was addressed as part of the conceptual model pathway between nutrients and macroinvertebrates. NMED has numeric criteria for minimum DO and pH in streams (5-6 mg/L DO, depending on the use; pH in the range of 6.6 – 9.0). This report examined DO and pH in correlation analyses and used DO in deriving TN and TP thresholds. pH was found to be positively correlated with both nutrients, more strongly with TN than with TP (see Table 15).

With increased nutrients there was greater daily fluctuations in DO and lower minimum DO (see Table 19). Minimum DO was negatively correlated to both TN and TP. Daily fluctuations (DeltaDO, Pmax4hr, and GPP) were positively and more strongly correlated to TP than to TN. DeltaDO, Pmax4hr, and Rmax4hr were highly correlated to each other.

In correlation analyses with macroinvertebrates, the diel DO metrics with the most significant correlations were minimum DO, Pmax4hr, and DeltaDO. Of these three metrics, Pmax4hr had the second greatest number of significant correlation in all sites and had the most significant correlation in each of the site classes. In the TP Flat-moderate site class, GPP had more significant correlations with macroinvertebrate metrics than other diel DO measure, though it was not highly correlated in all sites or in other site classes. NMED could continue to use minimum DO in assessments because it is correlated with both nutrients and macroinvertebrate metrics and is an important linkage in the conceptual model. Either Pmax4hr or DeltaDO could also be used for assessments because they are highly correlated with each other and with macroinvertebrate metrics. These DO metrics are correlated to TP more than they are to TN.

When minimum DO and Pmax4hr were used in change-point analysis to find nutrient thresholds, the identified thresholds were similar to the reference 90^{th} quantile values. This corroboration of nutrient thresholds and macroinvertebrate responses suggests that the DO metrics are appropriate indicators for NMED assessment protocols. Change-point analysis to find DO metric thresholds from macroinvertebrate metrics and nutrient concentrations resulted in median minimum DO thresholds of 2.3 mg/L, lower than the 5-6 mg/L currently in NMED standards. Only change-points identified from TN were in the 5-6 mg/L range.

Application Issues

The NMED SWQB might decide to select interim nutrient thresholds based on these analyses and ranges of threshold values. Selection of interim thresholds should be based on the ranges of potential thresholds derived from reference distributions and stressor-response analyses. In general, the thresholds derived from the 90^{th} quantile of reference distributions are higher than the median of those derived through stressor-response analyses. Selection of an interim threshold should weigh the merits of each analytical technique and the levels of protection they afford. The reference distribution approach emphasizes observed best conditions. Stressor-response approaches emphasize changes in macroinvertebrate and diatom metrics. NMED should establish the final thresholds using an adaptive management framework in which after testing preliminary thresholds with new data and scrutinized them in the context of appropriate policies and applications.

The thresholds derived in this report are based on data from perennial wadeable streams as defined by NMED. Any interim thresholds should only be applied in similar streams types.

7.0 Literature Cited

Antweiler, R.C. and H.E. Taylor. 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environ. Sci. Technol. 42:3732–3738.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Chapter 6 in the EPA Rapid Bioassessment Protocol for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, Second Edition, EPA 841-B-99-002.

Biggs, B.J.F. and C. Kilroy. 2000. Stream periphyton monitoring manual. Prepared for the New Zealand Ministry for the Environment. NIWA, Christchurch, NZ.

Biggs, B.J.F., 2000, Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae: Journal of American Benthological Society, v. 19, no. 1, p. 50-67.

Cade, B. S., J. W. Terrell, and R. L. Schroeder. 1999. Estimating effects of limiting factors with regression quantiles. *Ecology* 80(1):311-323

Calow, P. and L.J. Calow. 1975. Cellulase activity and niche separation in freshwater gastropods. Nature. 255:478-480.

Chetelat, J., F.R. Pick, A. Morin, and P.B. Hamilton. 1999 Periphyton biomass and community composition in rivers of different nutrient status. Canadian Journal of Fisheries and Aquatic Sciences 56:560-569.

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(1):17-23.

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

Dodds, W. K., V. H, Smith and B. Zander. 1997. Developing nutrient targets to control benthic chlorophyll levels in streams: a case study of the Clark Fork River. Water Research, 31(7), 1738-1750.

Dodds, W. K. and E. B. Welch. 2000. Establishing nutrient criteria in streams. J. N. Am. Benthol. Soc.19:186-196.

Downes, B. J., P. S. Lake, E. S. G. Schreiber, and A. Glaister. 2000. Habitat structure, resources and diversity: the separate effects of surface roughness and macroalgae on stream invertebrates. *Oecologia*, *123*(4), 569-581.

Griffith, G.E., J.M. Omernik, M.M. McGraw, G.Z. Jacobi, C.M. Canavan, T.S. Schrader, D. Mercer, R. Hill, and B.C. Moran. 2006. Ecoregions of New Mexico (color poster with map, descriptive text, summary tables, and photographs): Reston, Virginia, U.S. Geological Survey (map scale 1:1,400,000).

Hawkins, C. P., R. H. Norris, J. Gerritsen, R. M. Hughes, S. K. Jackson, R. K. Johnson, and R. J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society, 19(3), 541-556.

Helsel, D., 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley, New York.

Helsel, D. 2010. Much ado about next to nothing: Incorporating nondetects in science. Ann. Occup. Hyg., 54(3): 257–262.

Hill, W.R. and S.E. Fanta. 2008. Phosphorus and light colimit periphyton growth at subsaturating irradiances. Freshwater Biology 53:215-225.

Hill, W.R., S.E. Fanta and B.J. Roberts. 2009. Quantifying phosphorus and light effects in stream algae. *Limnology and Oceanography* 54:368-380.

Jacobi, G.Z., M.D. Jacobi, M.T. Barbour, and E.W. Leppo. 2006. Benthic macroinvertebrate stream condition indices for New Mexico wadeable streams. Prepared for the New Mexico Environment Department, Santa Fe.

Jessup, B.K., D. Eib, L. Guevara, J. Hogan, F. John, S. Joseph, P. Kaufmann, and A. Kosfiszer. 2010. Sediment in New Mexico Streams: Existing Conditions and Potential Benchmarks. Prepared for the U.S. Enviromental Protection Agency, Region 6, Dallas, TX and the New Mexico Environment Department, Santa Fe, NM. Prepared by Tetra Tech, Inc., Montpelier, VT.

Jessup, B.K., P.R. Kaufmann, F. John, L.S. Guevara, S. Joseph. 2014. Bedded sediment conditions and macroinvertebrate responses in New Mexico streams: a first step in establishing sediment criteria. Journal of the American Water Resources Association. *50*(6):1558-1574.

Jin, S., L.Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment*, 132: 159 – 175.

Justus, B.G., J.C. Petersen, S.R. Femmer, J.V. Davis, and J.E. Wallace. 2010. A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams. *Ecological Indicators* 10:627-638.

Kaufmann, P.R., D.V. Peck, and S.G. Paulsen. 2012. NRSA Technical Support Documentation; Fluvial (River and Stream) Physical Habitat Condition Assessment (DRAFT 7/16/2012). Prepared by the USEPAOffice of Research and Development, Corvallis, OR.

Kelly, M.G., and Whitton, B.A., 1995, The Trophic Diatom Index—A new index for monitoring eutrophication in rivers. Journal of Applied Phycology v. 7, p. 433–444.

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(6):795-809.

Lodge, D. M. 1991. Herbivory on freshwater macrophytes. *Aquatic Botany*, *41*(1):195-224.

Mabe, J.A., 2007, Nutrient and biological conditions of selected small streams in the Edwards Plateau, Central Texas, 2005–06, and implications for development of nutrient criteria: U.S. Geological Survey Scientific Investigations Report 2007–5195, 46 p.

Marks, J. C., Power, M. E. and Parker, M. S. 2000. Flood disturbance, algal productivity, and interannual variation in food chain length OIKOS 90: 20–27. Copenhagen

Mulholland, P. J., C.S. Fellows, J. L. Tank, N. B. Grimm, J. R. Webster, S. K. Hamilton, E. Martí, L. Ashkenas, W. B. Bowden, W. K. Dodds, W. H. McDowell, M. J. Paul and B. J. Peterson. 2001. Inter-biome comparison of factors controlling stream metabolism. *Freshwater Biology* 46(11):1503-1517.

Mulholland, P.J., J.N. Houser and K.O. Maloney. 2005. Stream diel dissolved oxygen profiles as indicators of in-stream metabolism and disturbance effects: Fort Benning as a case study. *Ecological Indicators* 5(3):243-252.

NMAC (New Mexico Administrative Code). 2005. State of New Mexico Standards for Interstate and Intrastate Streams. 20.6.4. New Mexico Water Quality Control Commission. As amended through July 17, 2005.

New Mexico Environment Department Surface Water Quality Bureau (NMED/SWQB). 2005. State of New Mexico Surface Water Quality Bureau Standard Operating Procedures for Sample Collection and Handling. Available at: http://www.nmenv.state.nm.us/swqb/. Santa Fe, NM.

________. 2008. State of New Mexico Nutrient Criteria Development Plan, Revision 4. Prepared by Surface Water Quality Bureau, New Mexico Environment Department. Accessed 01/04/2013 at: http://www.nmenv.state.nm.us/swqb/Nutrients/index.html

________. 2011. Procedures for Assessing Water Quality Standards Attainment for the State of New Mexico CWA §303(D) /§305(B) Integrated Report. Prepared by Surface Water Quality Bureau, New Mexico Environment Department. Accessed 01/04/2015 at: http://www.nmenv.state.nm.us/swqb/documents/swqbdocs/MAS/Protocols/AssessmentProtocol+Appendices-2010.pdf

________. 2012. State of New Mexico Surface Water Quality Bureau Standard Operating Procedure for Periphyton Sampling. Santa Fe, NM.

________. 2013. Procedures for Assessing Water Quality Standards Attainment for the State Of New Mexico CWA §303(d) /§305(b) Integrated Report: Nutrient Assessment Protocol for Wadeable, Perennial Streams. Surface Water Quality Bureau, New Mexico Environment Department.

________. 2014. Standard Operating Procedure (11.2) for Sampling Periphyton. Surface Water Quality Bureau, New Mexico Environment Department.

Omernik, J. M. 2006. Level III and IV Ecoregions of New Mexico (Version 1). US Environmental Protection Agency, Washington, DC.

Paul, M.J. 2008. Draft Predictive Bioassessment Models for New Mexico Streams. Prepared for New Mexico Environment Department, Santa Fe, NM. Prepared by Tetra Tech, Inc., Owings Mills, MD.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, S.A. Peterson, P.L. Ringold, T. Magee, and M. Cappaert, 2006. Environmental Monitoring and Assessment Program-Surface Waters Western Pilot Study: Field Operations Manual for Wadeable Streams. EPA ⁄ 620 ⁄R-06 ⁄ 003. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

Peterson, C.G., A.C. Weibel, N.B. Grimm, and S.G. Fisher. 1994. Mechanisms of benthic algal recovery following spates: comparison of simulated and natural events. Oecologia August 1994, Volume 98, Issue 3-4, pp 280-290

Porter, S. D. 2008. Algal attributes: an autecological classification of algal taxa collected by the National Water-Quality Assessment Program. US Geological Survey.

Potapova, M., and D.F. Charles, 2007, Diatom metrics for monitoring eutrophication in rivers of the United States: Ecological Indicators, v. 7, p. 48–70.

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(1):87-97.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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.

Stevenson, R. J, Y. Pan, K. M. Manoylov, C. A. Parker, D. P. Larsen, and A.T. Herlihy. 2008. Development of diatom indicators of ecological conditions for streams of the western US. *J. N. Am. Benthol. Soc*., 27(4):1000–1016.

Stoddard, J.L., D.V. Peck, A.R. Olsen, D.P. Larsen, J. Van Sickle, C.P. Hawkins, R.M. Hughes, T.R. Whittier, G. Lomnicky, A.T. Herlihy, P.R. Kaufmann, S.A. Peterson, P.L. Ringold, S.G. Paulsen, and R. Blair. 2005. Western Streams and Rivers Statistical Summary. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA 620/R-05/006.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., and Norris, R.H. 2006. Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition. Ecological Applications, 16(4), 2006, pp. 1267–1276.

Suplee, M.W., V. Watson, M. Teply, and H. McKee. 2009. How Green is Too Green? Public Opinion of What Constitutes Undesirable Algae Levels in Streams1. *JAWRA Journal of the American Water Resources Association*, *45*(1), 123-140.

Tetra Tech. 2014. Quality Assurance Project Plan for Nutrient-Scientific Technical Exchange Partnership System (N-STEPS) – Secondary Data Analysis and Model Development in Support of Numeric Nutrient Criteria. Revision 1. Prepared for U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Health and Ecological Criteria Division by Tetra Tech, Inc., Fairfax, VA.

________. 2012. Quality Assurance Project Plan for New Mexico Nutrient Framework Technical Support; QAPP 343. Prepared by Tetra Tech, Inc., Fairfax, VA, September 2012.

________. 2011a. Quality Assurance Project Plan for Nutrient-Scientific Technical Exchange Partnership System (N-STEPS) – Secondary Data Analysis. Revision 0. Prepared for U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Health and Ecological Criteria Division by Tetra Tech, Inc., Fairfax, VA.

________. 2011b. New Mexico Nutrients: Translator Development Approach and Proof of Concept. Prepared for: U.S. Environmental Protection Agency Region 6, Dallas TX and New Mexico Environment Department, Santa Fe, NM. 55 p.

Therneau, T., B. Atkinson and B. Ripley. 2013. rpart: Recursive Partitioning. R package version 4.1-3. http://CRAN.R-project.org/package=rpart

USEPA (U.S. Environmental Protection Agency). 2013. Biological Assessment Program Review: Assessing Level of Technical Rigor to Support Water Quality Management. EPA 820-R-13-001. Office of Science and Technology, Washington, DC 20460.

________. 2012. CADDIS website: Nitrogen & Phosphorus: Simple Conceptual Diagram. Accessed 11/17/2012 at http://www.epa.gov/caddis/ssr_nut4s.html

________. 2010. Using Stressor-response Relationships to Derive Numeric Nutrient Criteria. EPA-820-S-10-001.

________. 2009. Empirical Approaches for Nutrient Criteria Derivation. Science Advisory Board Review Draft.

________. 2007. National Rivers and Streams Assessment; Field Operations Manual, EPA-841-B-07-009, U.S. Environmental Protection Agency, Washington, D.C.

________. 2006. Estimation and Application of Macroinvertebrate Tolerance Values. Report No. EPA/600/P04/116F. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

________. 2004. Wadeable Stream Assessment: Field Operations Manual. EPA841-B-04-004. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.

________. 2000a. Nutrient Criteria Technical Guidance Manual: Rivers and Streams. EPA-822-B-00-002. U.S. Environmental Protection Agency, Washington, DC. http://www.epa.gov/ost/criteria/nutrient/guidance/rivers/index.html

________. 2000b. Ecoregional Nutrient Criteria Documents for Rivers & Streams. EPA 822-B-01-013, 015, and 016.http://www.epa.gov/waterscience/criteria/nutrient/ecoregions/rivers/index.html

________. 1998. National Strategy for the Development of Regional Nutrient Criteria. EPA/822-R-98-002. June 1998.

van Dam, H., Mertens, A., and Sinkeldam, J., 1994, A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands: Netherlands Journal of Aquatic Ecology, v. 28, no. 1, p. 117–133.

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 Sciences, 53(1), 99-105.

Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. Journal of Agricultural, Biological, and Environmental Statistics, 370-388.

Van Sickle, J., and R. M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society, 19(3), 370-384.

Wallace, J. B., and J.R. Webster. 1996. The role of macroinvertebrates in stream ecosystem function. *Annual review of entomology*, *41*(1), 115-139.

Woodruff, L., W.F. Cannon, D.B. Smith, and F. Solano, 2015. The distribution of selected elements and minerals in soil of the conterminous United States, J. Geochem. Explor., http://dx.doi.org/10.1016/j.gexplo.2015.01.006)