UTAH DEPARTMENT *of* ENVIRONMENTAL QUALITY
WATER QUALITY

TECHNICAL SUPPORT DOCUMENT: UTAH'S NUTRIENT STRATEGY

Scientific Investigations to Support Utah's Nutrient Reduction Program

June 2019

Prepared by the Utah Division of Water Quality

TECHNICAL SUPPORT DOCUMENT: UTAH'S NUTRIENT STRATEGY

Scientific Investigations to Support Utah's Nutrient Reduction Program

2019

Authors

Prepared by the Utah Division of Water Quality

J. D. Ostermiller
M. Shupryt
M. A. Baker
B. Neilson[2]
E. B. Gaddis
A J. Hobson[2]
B. Marshall[1]
T. Miller[1]
D. Richards[1]
N. von Stackelberg

1. Primary authors for Chapter 15
2. Primary authors for Chapter 16

Technical Review

T. Bosteels
E. Flemer
T. Laidlaw
D. Sorenson
C. Walker

ACKNOWLEDGMENTS

The authors wish to thank the many individuals who assisted with this research. Particularly Utah's Water Quality Board who provided project funding. The Division of Water Quality's Monitoring Section did a miraculous job collecting the information for this report; their many long hours and devotion to the project are greatly appreciated. The Nutrients Workgroup within the Division of Water Quality, the Nutrient Core Team, and the Nutrient Technical Review Team provided valuable input into preliminary results.

EXECUTIVE SUMMARY

> The primary objective of this report is to provide the technical basis for the development of numeric nutrient criteria for Utah's rivers and streams.

Introduction

Worldwide, humans continue to add excessive levels of nutrients, particularly nitrogen (N) and phosphorus (P), to waterbodies. These inputs have created what many, including Utah's Division of Water Quality (DWQ) and the U.S. Environmental Protection Agency (USEPA), consider to be among the most significant threats to water quality. The response to these concerns has been a nationwide effort to reduce human-caused inputs of nutrients to waterbodies. One important component of efforts to minimize the deleterious consequences of nutrient enrichment on streams and lakes is the development of numeric nutrient criteria (NNC), which define N and P concentrations that cannot be exceeded if ongoing support of the beneficial uses of these waterbodies is to be maintained.

Nutrients act through many interrelated paths that can lead to the degradation of aquatic life, drinking water, or recreation uses (Figure ES.1). These paths between nutrients and uses are not independent. They are moderated by locally divergent physical and chemical processes. From a technical basis, the complexity of these relationships is a central challenge in addressing nutrient-related water quality problems. Other challenges are socioeconomic. There are many different sources of nutrients and, therefore, a diverse set of stakeholders who often have conflicting interests. Additionally, reducing nutrient inputs to streams and lakes can be costly.

To address the complexity of cultural eutrophication problems DWQ has adapted an adaptive management framework. This approach encourages iterative solutions that can be applied while uncertainties that result from the complexity of the problems are resolved. In Utah, numerous streams and lakes have been listed as impaired due to violations of water quality parameters associated with excess nutrients (e.g., Dissolved Oxygen, pH, trophic state indexes). In response to these impairments, several total maximum daily load nutrient management plans have been developed for watersheds throughout the state. These reactive approaches are insufficient because they too frequently rely on impairment to occur before action is taken. As a result, DWQ, in collaboration with key stakeholders (the Nutrient Core Team), has proposed initial solutions that seek affordable reductions from both point- and nonpoint sources. In addition, the Nutrient Core Team proposes an iterative development of regulations beginning with the promulgation of NNC in headwater streams, to be followed by site-specific criteria development for streams elsewhere. Chapter 1 provides additional details about Utah's iterative approach to NNC development.

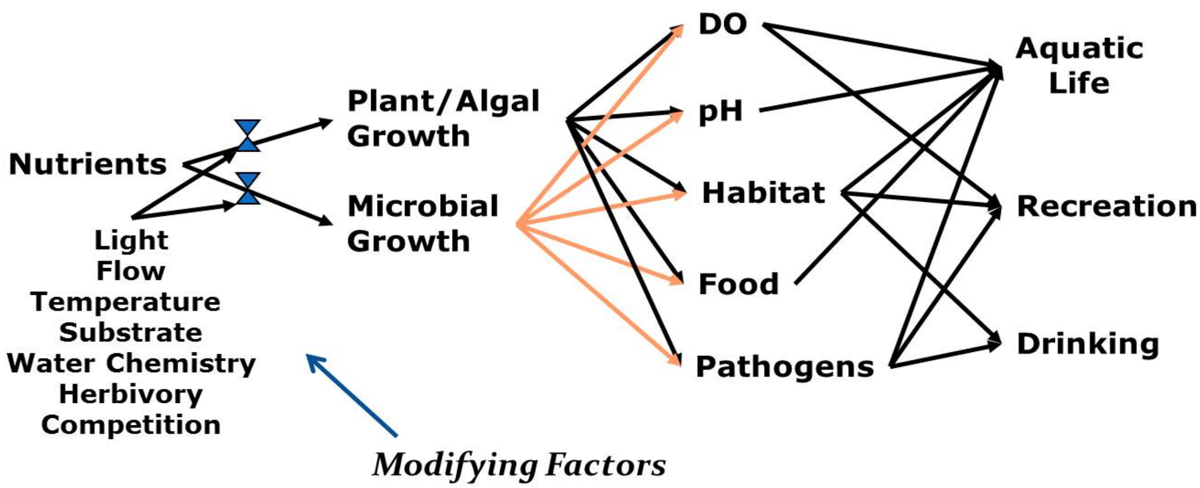


Figure ES.1. Simplified model depicting linkages between nutrients and common designated uses in streams (source TetraTech).

The primary objective of the research described in this technical support document (TSD) is to provide the technical basis for the development of NNC to protect aquatic life and recreation uses (Section 1). Several ecological responses known to be sensitive to nutrient enrichment and relevant to the protection of aquatic life uses were identified from the scientific literature. These responses fall within two broad classes: (1) functional responses that can be used to quantify changes in important processes or states, and 2) structural responses that quantify changes in the composition and abundance of stream biota. In both cases, Utah-specific thresholds were empirically derived to determine concentrations of N or P that are broadly associated with stream condition as measured by each response. Two thresholds were derived for each indicator: one to distinguish between streams in good versus fair condition and one to distinguish between streams in fair versus poor condition. To protect recreation uses, a survey was conducted that examined the influence of excess algae growth on stream recreation activities (Chapter 8). Once established, the ecological relevance of these statistical thresholds was evaluated with independent measures of stream condition and further corroborated by comparing the results of these studies with those published in primary scientific literature. Figure ES.2 summarizes all aquatic life thresholds obtained from these inquiries.

The intended application of the ecological response thresholds is place dependent. For headwater streams, the empirical thresholds described in the first two sections of the report provide multiple lines of evidence that DWQ used to propose NNC to protect these waters (Section 3). For streams outside of headwaters the stressor-response relationships will be used to inform assessment methods so that DWQ can better identify those waterbodies with nutrient-related problems. As streams with nutrient-related problems are identified, the responses described in this TSD will provide useful diagnostic information because each indicator describes different aspects of the myriad of possible ecological responses to excess nutrients. These diagnostic data will help inform study designs for the derivation of site-specific standards and specific remediation actions that are most likely to restore stream conditions.

Ecological Responses

Different approaches were required for the derivation of functional and structural response thresholds. Existing data were not available for functional indicators, so DWQ established a pilot study to generate measures of these responses from 15 reference condition streams and 20 additional streams that varied in extent of nutrient enrichment (Chapter 3 provides study design details). In the case of structural responses, DWQ has an established biological assessment program, so existing sources of data could be used to determine the concentrations of N or P that were associated with the greatest changes in the composition of macroinvertebrates and diatoms (Chapter 7). While these indicators do not capture all possible responses, collectively they provide quantitative measures of all important causal paths between nutrients and responses. The selected indicators are pragmatic because they either capitalize on responses derived from routine and ongoing monitoring or depend on data that could be collected as part of ongoing monitoring efforts with minimal additional resources.

Functional Responses

Nutrient-diffusing substrates (NDS; Chapter 4) were used to provide quantitative estimates of two important states: the nutrient that primarily limits algae production and the concentration of N or P that is associated with saturation—the point where additional nutrient increases do not result in increases in primary production. NDSs are bioassays where growth media are augmented with N, P, or N and P. These experiments are deployed in streams, and the accrual of algae on treatments is compared with controls. These investigations highlight the importance of considering both N and P in Utah's nutrient reduction strategy. Co-limitation of both N and P was the most common condition among study streams, a finding that is consistent with recent literature. Among streams that were not saturated with nutrients, the addition of both N and P in these bioassays resulted in a greater response than the addition of either nutrient alone, which suggests that in streams with excess algae growth the reduction of both N and P is more likely to improve conditions than reductions of either nutrient alone. Thresholds calculated with these data revealed a total P (TP) saturation threshold of 0.078 mg/L (± 0.017–1.33), and a total N (TN) saturation threshold of 0.42 mg/L (± 0.33–1.4) (Figure ES.2).

Whole stream metabolism techniques were used to obtain measures of two fundamental ecological functions: gross primary production (GPP), which measures rates of primary production via concurrent oxygen production, and ecosystem respiration (ER), which measures the growth of animals as utilization of carbon and the concurrent consumption of oxygen by animal and microbes (Chapter 5). Thresholds were established for TN and TP concentrations that were associated with streams classified as being in good, fair, or poor condition based on measured GPP and ER rates. These calculations suggest that, on average, streams move from good to fair condition at a TP above 0.02 mg/L or a TN of 0.09 mg/L. Similarly, streams generally move from fair to poor condition once TP exceeds 0.09 mg/L or TN exceeds 1.28 mg/L (Figure ES.2). Thresholds were also calculated for GPP and ER so that these metrics could be used to assess stream conditions. In general, streams are in poor condition once GPP exceeds 10 g O_2/m^2/day or ER exceeds 9 g O_2/m^2/day; however, the confidence is this assertion is lower for streams

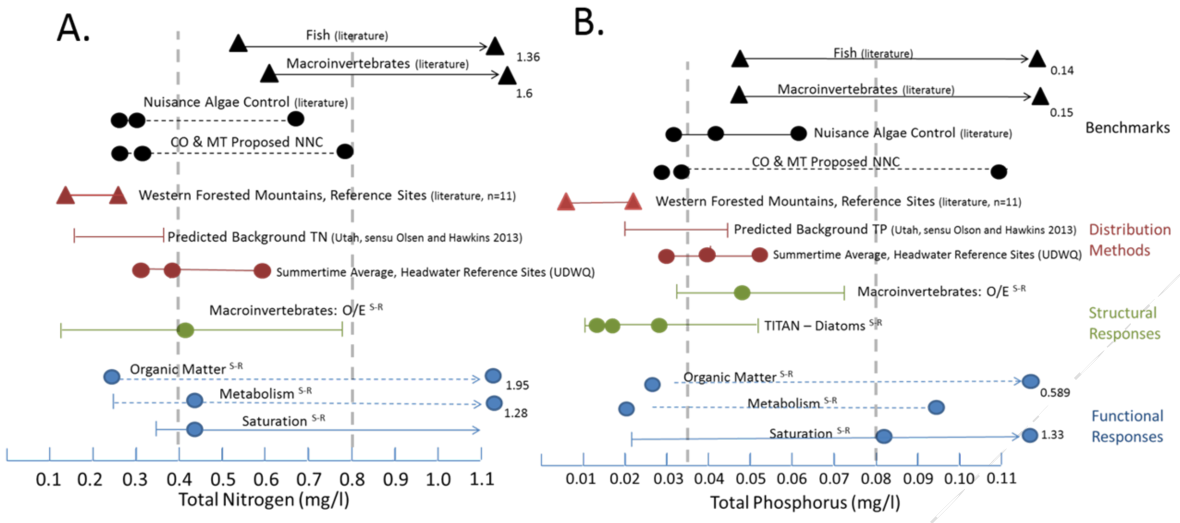


Figure ES.2. Numeric nutrient criteria thresholds derived from numerous sources for total nitrogen (panel A) and total phosphorus (panel B), along with the proposed numeric nutrient criteria for these nutrients presented in this technical support document.

Notes: Lines bracketed by triangles indicate the omission of numerous intermediate thresholds (dots). The graphics are colored to demarcate different categories of thresholds. Blue denotes functional responses. Green denotes structural responses. Red denotes thresholds derived using frequency distribution methods; the bottom red dots indicate the 75th, 90th, and 95th percentiles of the summertime average of Utah reference sites, the middle red line denotes background concentrations obtained from an empirical model that predicts background concentrations from natural environmental gradients, and the top red line denotes other distribution methods from reference site distributions in USEPA Nutrient Ecoregion II (Evans-White et al. 2014). Black denotes broad benchmarks for other proposed numeric criteria from USEPA Region 8 (the bottom black line) and values obtained from primary literature (the top three black lines). The vertical dotted lines are the proposed numeric nutrient criteria thresholds presented in this technical support document.

with low slopes or high canopy cover. Metabolism response thresholds were compared against independently derived numeric criteria for dissolved oxygen (DO). Streams with DO observations below numeric criteria occurred at sites with high GPP and ER. The empirically derived thresholds for GPP and ER from Utah streams were consistent with deleterious responses reported in the primary literature.

The amount of storage and processing of carbon is another important ecosystem function of direct relevance to cultural eutrophication. Excessive GPP can lead to excessive algae growth that degrades aquatic life uses via diminished quality of habitat or food or recreation uses via diminished aesthetics. The consumption (i.e., ER) of excessive carbon within a system—from plant and algae growth and outside sources—can result in low DO, particularly when ambient N or P is high. This study captured the importance of carbon with reach-scale measures of organic matter standing stocks in the stream (Chapter 6). Thresholds for TN and TP were derived for sources of carbon within the stream that are most strongly related to primary production or most readily available to stream fungus and microbes (e.g., fine particulate organic matter). Streams with high levels of these sources of organic matter also had higher concentrations of TN and TP. Statistical relationships further suggest that when stream TP exceeds 0.026 mg/L or TN exceeds 0.238 mg/L sites move from good to fair condition, and concentrations of 0.589 and 1.95 mg/L (TP and TN, respectively) distinguish between streams in fair versus poor condition (Figure ES.2).

Organic matter standing stock thresholds were also calculated, which suggested that streams with stores of more than 48 g ash free dry mass/m^2 are at increased risk of having degrading uses. These thresholds were confirmed against independently derived DO criteria, which indicated that sites where standing stocks were above these thresholds were much more likely to have DO observations below criteria.

Structural Responses

Excess nutrients alter ecological functions, but it is important to know the extent to which these changes alter stream biota. The functional indicator pilot data were augmented with existing biological assessment data to evaluate relationships between increasing stream nutrients and alterations to the presence and relative abundance of macroinvertebrate and diatom assemblages using a statistical technique called threshold indicator taxon analysis (TITAN; Chapter 7). Historic laboratory procedures limited the analysis of both assemblages to TP because TN was available at an insufficient number of sites with diatom taxa to conduct the analyses. TITAN uses changes in the presence/absence and relative abundance of taxa within each assemblage to derive three thresholds: one that identifies the TN or TP concentration that best identifies losses of sensitive taxa, one that identifies concentrations associated with increases in tolerant taxa, and an overall threshold that identifies the concentration where both groups exhibit the strongest changes. All three thresholds were calculated and reported, but this summary is limited to the third (average) thresholds. For TP, a threshold of ~0.02 mg/L was established for both diatoms (0.022 mg/L ± 0.010–0.047) and macroinvertebrates (0.015 mg/L ± 0.004–0.113); confidence intervals given are the 5^{th} and 95^{th} intervals. A TN threshold of 0.41 mg/L (± 0.40–1.1) identified the concentration that, on average, was associated with the greatest changes in the composition of macroinvertebrates. For macroinvertebrates, thresholds for TN and TP were compared against independently derived biological impairment thresholds (the ratio of the number of observed species compared to the number of expected species [O/E] > 0.78 or 0.83). On average, streams that were predetermined to be impaired from biological assessments had TN or TP concentrations above the thresholds established for TITAN. The thresholds for both diatoms and macroinvertebrates were also consistent, albeit somewhat lower, than those obtained from similar evaluations elsewhere (Figure ES.2).

Recreation Responses

Excessive nutrients can also degrade recreation uses. In some cases, nutrient-related degradation of recreation uses is caused by increases in pathogens or biological toxins. More frequently, particularly for small- to moderate-size streams, recreation uses are potentially degraded by decreased aesthetics from

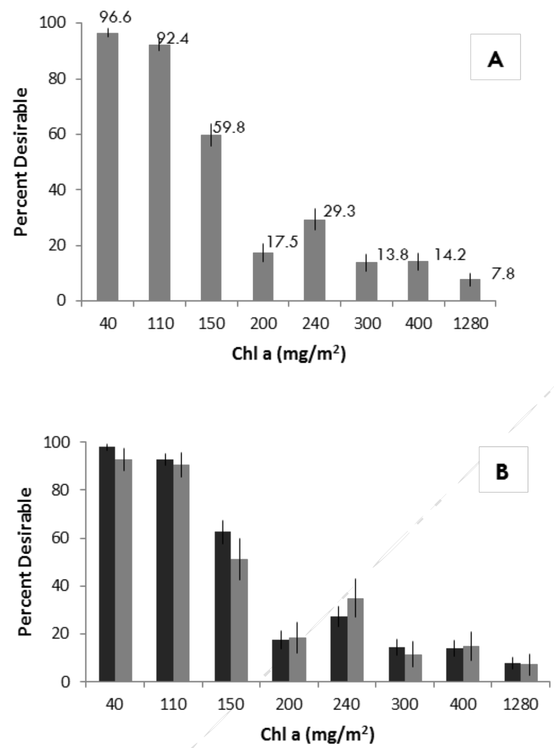


Figure ES.3. Percent of all survey respondents who indicated that streams of varying algae concentrations were reflective of desirable conditions (Panel A). Also depicted is the distinction between respondents who recreate at streams (Panel B, black bars) and those who do not (Panel B, gray bars).

excess benthic algae. To quantify the latter, an investigation was conducted to relate aesthetics, recreation, and algae growth (Chapter 8). Surveys that included pictures of streams with different concentrations of benthic algae were mailed to 2,700 randomly selected Utah households. For each randomly ordered picture, survey participants were asked whether they would consider the depicted conditions to be desirable or undesirable to their recreation experience. The majority of the 628 respondents indicated a shift from desirable to undesirable conditions as algae densities increased from 110 to 150 mg chlorophyll-*a* (chl-*a*)/m^2 (Figure ES.3). These opinions did not differ between citizens who reported recreating on streams (users) versus those who did not (nonusers).

Development of Numeric Nutrient Criteria for Utah's Headwater Streams

The second section of the report provides the technical basis for the application of the thresholds derived in Section 1. Section 2 of the report includes two chapters that use historical water quality data collected at headwater streams for two different, but related, purposes. The first was a classification exercise where ambient nutrient data were combined with measures of natural environmental gradients to determine whether headwater streams needed to be classified into smaller groups of streams to minimize natural variation in nutrient concentrations (Chapter 9). The second used ambient nutrient data collected at all headwater streams and at headwater streams designated as reference sites to analyze the distribution of TN and TP (Chapter 10). The distribution of TN and TP among both groups of streams formed one line of evidence for the proposed NNC thresholds.

A weight of evidence approach is used to bring together all the lines of evidence presented in Sections 1 and 2 to make recommendations for appropriately protective NNC. Because distillation of many different lines of evidence into the regulatory values incorporated into the NNC requires professional judgment, Chapter 11 attempts to explain the underlying rationale for how the proposed criteria were established. Headwater NNC are proposed that combine ambient nutrient concentrations with ecological responses to make determinations of aquatic life use support. Two nutrient thresholds are established that place streams into one of three bins based on the risk that nutrient enrichment poses to these uses. Streams where the average nutrient concentrations fall below 0.035 mg/L for TP *and* 0.40 mg/L for TN during the period of algae growth through senescence are considered indistinguishable from reference condition and in full support of aquatic life uses (with respect to nutrient pollution), unless other indications of deleterious responses to cultural eutrophication are observed. In contrast, streams where TP is > 0.080 *or* TN > 0.80 mg/L are considered to be impaired. For streams where the concentrations of TN and TP falls between these thresholds, ecological confirmation is required to determine whether nutrient enrichment has caused harm to aquatic life. Because nutrient enrichment can degrade uses through both autotrophic and heterotrophic pathways, the selected indicators cover both. For autotrophic responses, the proposed NNC state that GPP cannot exceed 6 g O_2/m^2/day and filamentous algae cover cannot exceed one-third of the stream bed. For heterotrophic responses, the proposed NNC specify that ER cannot exceed 5 g O_2/m^2/day.

After proposing the headwater NNC an independent study was conducted in 2015 at 49 headwater streams (Chapter 13). The results of this study provide further support for the proposed NNC and generally confirm that previous efforts to protect water quality in these ecosystems have been successful.

Ongoing Implementation of Nutrient Reduction Efforts

The final section of the report covers several important considerations for the interpretation of the headwater NNC and for the development of site-specific NNC for streams lower in the watersheds.

Regional stressor-response (S-R) patterns provide insights that can inform management objectives, but they have limitations. In particular, these approaches are unable to capture the influence of covariates on S-R relationships, particularly those that vary across local, site-specific scales. Accordingly, DWQ has opted to generate site-specific NNC for streams outside of headwaters for those streams where their development has been prioritized. This decision provides an opportunity to address several sources of uncertainty in S-R relationships that are very difficult to address at a regional scale. Specifically, site-specific investigations offer an opportunity to embrace the intrinsic complexity of streams, to reduce uncertainty, and to strengthen causal S-R inferences. These site-specific investigations also provide the opportunity to evaluate the relative influence of nutrients and other stressors on ecological responses. The potential power of these site-specific investigations requires a carefully crafted study design, which involves consideration of as many potential sources of variation as possible. Specific guidance with regard to these considerations is provided in Chapter 15.

In accordance with adaptive management principles, the work described in this report represents one step among many in the ongoing development of Utah's nutrient reduction strategy. The S-R relationships developed in this report will continue to be refined as additional investigations continue. New indicators will likely be developed. These improvements will continue to improve the accuracy of nutrient-related assessments and the defensibility of site-specific NNC. Practically speaking, these tasks—among others—are best achieved by embracing principles of collaborative management. Many have a direct interest in setting appropriately protective NNC. The reduction of scientific uncertainty hinges upon translating these collective interests into a shared understanding of the nutrient-related problems and appropriate regulatory responses.

CONTENTS

LIST OF FIGURES

FIGURES IN APPENDICES

LIST OF TABLES

TABLES IN APPENDICES

ABBREVIATIONS AND ACRONYMS

Acronym/ Abbreviation	Description
ABL	Aquatic Biogeochemistry Laboratory at Utah State University
AFDM	Ash free dry mass
ANOVA	Analysis of variance
AUC	Area under the curve
awch	Available water holding capacity of soils
BACI	Before, after, control, impact study design
BMP	Best management practice
BOD	Biological oxygen demand
CADDIS	Causal Analysis/Diagnosis Decision Information System
CBOD	Carbonaceous biochemical oxygen demand
CBOM	Coarse benthic organic matter
chl-*a*	Chlorophyll-*a*
CI	Confidence interval
COV	Coefficients of variation
CPOM	Coarse particulate organic matter
CWA	Clean Water Act
DEM	Digital elevation model
DEQ	Department of Environmental Quality
DIN	Dissolved inorganic nitrogen
DO	Dissolved oxygen
DWQ	Utah Division of Water Quality
Elev	Elevation at the bottom of the watershed
ER	Ecosystem respiration
FBOM	Fine benthic organic matter
FDM	Frequency distribution methods
GLM	Generalized linear models
GPP	Gross primary production

Acronym/ Abbreviation	Description
GSL	Great Salt Lake
HAB	Harmful algae blooms
HSD	Honestly significant difference
HU	Habitat unit
HUC	Hydrologic unit code, a cataloging code used to distinguish watersheds
ISS	Inorganic suspended solids
K-M	Kaplan-Meier (K-M) method of survival analysis
kfact	Soil erodability factor
MDL	Method detection limits
MEANP	Mean annual precipitation
MSE	Mean squared error
N	Nitrogen, expressed in mg/L throughout this report
NAWQA	National Water-Quality Assessment
nCPA	Nonparametric change point analysis
NDR	Nonparametric deviance reduction
NDS	Nutrient-diffusing substrates
NHST	Null hypothesis significance testing
NNC	Numeric nutrient criteria
NRSA	National Rivers and Streams Assessment
NTU	Nephelometric turbidity units
O/E	A biological assessment metric that compares the number of taxa observed at a site, O, to those that were predicted (expected) to occur without human disturbance, E
OM	Organic matter
P	Phosphorus expressed in mg/L throughout the report
PAR	Photosynthetic active radiation
PCA	Principal components analysis
perm	Water permeability of soils
POM	Particulate organic matter

Acronym/ Abbreviation	Description
POTW	Publicly owned treatment works
PP	Prediction probability
PRISM	Physiographically sensitive mapping of temperature and precipitation
QAPP	Quality assurance project plan
RCC	River continuum concept
RIVPACS	River invertebrate prediction and classification system
RMSE	Root mean square error
ROC	Receiver operator characteristics
RPS	Recovery potential screening
RR	Relative risk
SAP	Sample analysis plan
sCBOD	Soluble carbonaceous biochemical oxygen demand
SOD	Sediment oxygen demand
SOP	Standard operating procedure
S-R	Stressor-response relationship, in this case nutrients-ecological responses
SRP	Soluble reactive phosphorus
STATSGO	State soil geographic database
STORET	STOrage and RETrieval water quality database maintained by USEPA
TIN	Total inorganic nitrogen
TITAN	Threshold indicator taxon analysis
TMDL	Total maximum daily load
TMEAN	Annual mean predicted air temperature
TN	Total nitrogen
TP	Total phosphorus, expressed in mg/L throughout the report
TSD	Technical support document
TSS	Total suspended solids
UAC	Utah Administrative Code
UCASE	Utah's Comprehensive Assessment of Stream Ecosystems
UDEQ	Utah Department of Environmental Quality

Acronym/ Abbreviation	Description
UPHL	Utah Unified Public Health Laboratories
USEPA	U.S. Environmental Protection Agency
USFS	U.S. Forest Service
USGS	U.S. Geologic Survey
USU	Utah State University
VSS	Volatile suspended solids
WLA	Wasteload allocation
WoE	Weight of evidence
WQBEL	Water quality based effluent limits
WRF	Water reclamation facility
WSA	Wadeable Streams Assessment
WT	Water tracing
WWTP	Wastewater treatment plant
XWD	Predicted number of days with precipitation

Chapter 1

BACKGROUND: AN INCREMENTAL APPROACH TOWARD NUMERIC NUTRIENT CRITERIA DEVELOPMENT

Key Points

Numeric nutrient criteria are important regulatory tools used in efforts to address water quality problems caused by cultural eutrophication.

Division of Water Quality has prioritized the development of numeric nutrient criteria for headwater streams because of their ecological and economic importance to Utahns.

Development of regional and site-specific numeric nutrient criteria for other types of waterbodies will be ongoing as nutrient-related water quality problems are identified.

Introduction

Since the Industrial Revolution, humans have altered important ecological characteristics of the biosphere (Lewis and Maslin 2015). Among these alterations has been an extensive modification of the biogeochemical cycles of nutrients, particularly carbon (C), nitrogen (N), and phosphorus (P) (Sietzinger et al. 2010). One ramification of these changes has been a significant increase in nutrients in waterbodies (Bouman et al. 2005). Worldwide, the rate of biologically available N and P delivered to the biosphere has more than doubled over the last 5 decades (Bennett et al. 2001, Galloway et al. 2004, Sietzner at al. 2010). The U.S. Environmental Protection Agency (USEPA) estimates that nearly half of the streams and rivers in the country now have moderate to high concentrations of N and P (USEPA 2006). These increases in nutrient inputs are directly attributable to the production of numerous goods and services needed to support the current exponential growth of the human population. A large portion of these nutrients are consumed by humans, but they are also recycled back to the biosphere and ultimately transported into aquatic ecosystems through a variety of sources (Table 1.1).

Table 1.1. Natural and human-caused sources of nutrients to aquatic ecosystems.

Natural	Human-Caused or Human-Altered
Atmospheric deposition Decomposition of plants and algae Erosion from nutrient-bearing rocks and soils Wildlife waste	Atmospheric deposition associated with fossil fuel consumption Discharges of treated wastewater Industrial discharges Overflow from combined storm and sanitary sewers Runoff from agricultural fields including irrigation return flow Runoff from pasture or range Septic tank leachate Stormwater

The increases in production and use of nutrients have benefitted humans in some respects. Production and use of nutrients have allowed agricultural production to largely keep pace with human population growth. However, these activities have also caused deleterious consequences, particularly when too many nutrients are delivered to aquatic ecosystems. Excessive human-caused delivery of nutrients to waterbodies, sometimes called cultural eutrophication, is causing aquatic ecosystems to degrade so significantly that some believe that nutrient enrichment is among the most important water quality problems worldwide (Smith 2003).

Cultural eutrophication can potentially diminish many of the benefits that people and other organisms obtain from aquatic ecosystems. Degradation of aquatic life, recreational uses, and drinking water quality due to nutrient enrichment is well documented. In the United States, approximately one-third of the streams and rivers are impaired due to excess levels of N or P (USEPA 2004). These problems are of particular concern to USEPA and states charged with upholding the Clean Water Act (CWA) because the support and maintenance of these uses is the central objective of the CWA. As a result, for over a decade, addressing cultural eutrophication has consistently been one of the highest water quality priorities, both nationally (USEPA 2011) and locally.

For the last several decades, worldwide acknowledgment of ongoing cultural eutrophication has led to considerable work focused on finding equitable solutions to the problem. One regulatory tool that has been central to these efforts is the development of numeric nutrient criteria (NNC) that would set clear limits on nutrient concentrations that should not be exceeded if the designated uses of waterbodies are to be maintained. NNC, once established, would become integral to many CWA regulatory programs such as helping to establish permit limits for point sources of N and P, identifying waterbodies with nutrient-related impairments, identifying restoration targets for waters where the deleterious effects of excess nutrients have been observed, and other programs. However, despite considerable effort to develop NNC, their adoption into state and federal water quality standards has been slow. Some of the barriers to NNC adoption relate to scientific uncertainty about the point at which nutrient enrichment is sufficient to harm uses. In addition, it is sometimes difficult to establish background nutrient concentrations due to the large number of human-caused and natural N and P sources. Other ongoing challenges are related to socioeconomic factors associated with the expense of nutrient reduction efforts and related impacts to public and private sectors of the economy.

The development of NNC remains important. However, the slow pace of NNC adoption, coupled with ongoing nutrient enrichment, has led to general acknowledgment that addressing cultural eutrophication requires a variety of regulatory and policy approaches. The widespread prevalence of human-caused nutrient enrichment, transport of nutrients to downstream waters from local sources, accumulation of nutrients downstream, and ongoing increases in human population will combine to increase the magnitude and extent of cultural eutrophication problems if reductions in N and P are not addressed through a wide range of nutrient reduction efforts.

In recent years, USEPA has acknowledged the need for a variety of solutions and has issued guidance that allows states the flexibility necessary to innovate and respond to local water quality needs (USEPA 2011). This policy includes elements of accountability to help ensure an ongoing reduction in cultural eutrophication problems. Specifically, states are required to develop a comprehensive nutrient reduction program that describes how nutrient reductions will be sought from all important sources, how watersheds will be prioritized for implementation of these reductions, and how a monitoring program to evaluate the relative success of ongoing nutrient reduction efforts will be implemented. States are also required to develop a plan for the ultimate adoption of NNC for all surface waters.

The State of Utah Division of Water Quality (DWQ) has long acknowledged the importance of cultural eutrophication and welcomes the flexibility in USEPA's guidance that allows for development of a nutrient reduction program tailored to the unique economic and environmental considerations in Utah. DWQ has identified numerous waters where cultural eutrophication has already resulted in degradation to aquatic life uses and has developed total maximum daily load (TMDL) requirements to address these problems. This previous work, combined with observations of long-term water quality trends, has led DWQ to conclude that equitable solutions to nutrient enrichment are needed to protect the quality of life for future generations. Utah's population is expected to double by 2050. This change in population will increase nutrient inputs unless policies are in place to reduce per capita nutrient inputs to the state's waters. Left unchecked, this nutrient enrichment is likely to exacerbate already strained water demands since degraded waters are less effective at meeting societal needs. For example, excessive algae growth resulting from excessive nutrients increases the cost of water treatment for culinary uses. One example is the formation of harmful algae blooms (HABs) that can produce potent natural toxins. The use of reservoirs affected by HABs as culinary water sources has been temporarily precluded in Utah and elsewhere. These blooms have also caused negative economic impacts to agriculture and other industries. The development of reasonable nutrient reduction policies is one way that DWQ can help ensure that these problems, among others, do not continue to worsen.

DWQ and EPA have made addressing nitrogen and phosphorus pollution a water quality priority. The agencies are addressing this growing water quality concern to ensure ongoing protection of the designated uses assigned to our lakes and rivers and to ensure the quality of life for future generations.

The State of Utah continues to use a collaborative approach to develop sensible nutrient reduction approaches for the state. For several years DWQ has been working with governmental and private sector leaders on a comprehensive nutrient reduction plan (www.nutrients.utah.gov). This "Nutrient Core Team" has recommended a multifaceted and iterative approach to nutrient reductions that balances the potential risks of ongoing enrichment against the costs associated with nutrient reductions from various sources. For example, effluent limits achievable with currently available and cost-effective technology (technology-based effluent limits) were established to keep P from municipal sources from increasing as the population increases. Another example is the establishment of an environmental stewardship certification to help reduce nutrient inputs from agricultural sources. DWQ has also implemented a policy that requires larger communities to include plans for reduction of stormwater nutrient sources in municipal separate storm sewer system permit applications. The common rationale for these recommendations was that nutrient reduction efforts would be relatively inexpensive to implement. Also, if successful, these early efforts to decease N and P pollution will help prevent cultural eutrophication water quality problems from worsening while DWQ carries out site-specific studies to obtain the data necessary to support NNC in others waters where NNC development is prioritized. The data requirements for the development of site-specific NNC are often resource intensive. However, because site-specific determinations of nutrient endpoints are generally more accurate than regional generalizations, these studies will provide detailed data to support and defend any future required reductions of nutrient inputs. DWQ and collaborators will use the site-specific NNC data to demonstrate the need for nutrient reduction efforts to protect downstream uses before additional, more costly, reductions are required.

Numeric Nutrient Criteria Development: Background Considerations

Regulatory Considerations

The CWA authorizes states and tribes to develop and adopt water quality standards to protect the chemical, physical, and biological integrity of their surface waters. Water quality standards consist of designated and beneficial uses, water quality criteria to protect these uses, and antidegradation policies that aim to protect high quality waters. States assign designated uses to waterbodies (e.g., recreation, agriculture, aquatic life) that define the goals for the waterbody. Water quality criteria—both narrative and numeric—define chemical or physical conditions that should not be exceeded to protect these uses. Numeric criteria (Utah Administrative Code [UAC] R317-2-14) provide specific pollutant concentrations

Utah's Narrative Criteria (UAC R317-2-7.2)

It shall be unlawful, and a violation of these regulations, for any person to discharge or place any waste or other substance in such a way as will be or may become offensive such as unnatural deposits, floating debris, oil, scum or other nuisances such as color, odor or taste; or cause conditions which produce undesirable aquatic life or which produce objectionable tastes in edible aquatic organisms; or result in concentrations or combinations of substances which produce undesirable physiological responses in desirable resident fish, or other desirable aquatic life, or undesirable human health effects, as determined by bioassay or other tests performed in accordance with standard procedures; or determined by biological assessment in Subsection UAC R317-2-7.3.

(e.g., magnitude) that must not be exceeded. Narrative criteria are more qualitative and describe conditions that should be avoided—or conversely those that must be maintained—for a waterbody's uses to remain protected. In contrast, Narrative criteria are intentionally broad, which has the advantage of capturing deleterious water quality conditions that are difficult to quantify. As a result, Utah's narrative criterion already addresses several deleterious nutrient effects and includes several conditions that are not permissible, such as: "scum," "undesirable aquatic life," or "objectionable taste or odor"; nevertheless, DWQ has found it difficult to develop effective regulations based exclusively on these qualitative goals. It is not always clear how to translate narrative statements to permit limits. Generally, narrative criteria require the development of quantitative guidelines before they can be integrated into water quality regulatory programs. If nutrient-related water quality problems are to be rectified and prevented, Utah needs a path that will lead to numeric N and P criteria.

Technical Considerations

NNC are derived on a site-specific basis or, for a broad class of waterbodies with similar characteristics, in a region of interest. Scientific investigations and subsequent analytical methods used to support site-specific and regional NNC can be similar, but the interpretation of the results can differ considerably depending on the spatial extent of their intended application. The USEPA recommends two main approaches for setting regional NNC (USEPA 2000): (1) frequency distribution methods (FDMs) that derive NNC thresholds based on percentiles of the distribution of ambient N and P concentrations, and (2) the use of empirical models that establish NNC thresholds from stressor-response (S-R) relationships. In contrast, site-specific approaches rely more heavily on mechanistic models that can more accurately account for the effects of naturally occurring physical and chemical stream characteristics on responses to nutrient enrichment. Investigations used to support regional and site-specific NNC produce data complementary to both approaches, provided that the information is properly interpreted.

This technical support document (TSD) presents the results of investigations that support the application of all three NNC development approaches. The methods used in these investigations are briefly introduced here to provide context for how each line of evidence informs the headwater NNC.

Stressor-Response Models

S-R models are the most frequently used analytical method to derive regional NNC. These empirical methods relate stressors (e.g., N or P) to ecological responses such as changes in biological composition (ecosystem structure) or biogeochemical processes (ecosystem functions). To establish these relationships, stressor and response data are obtained from numerous sites that encompass the range of conditions within a region of interest. Next, statistical models are used to establish nutrient-concentration thresholds for the nutrients that are most strongly associated with changes in ecological responses. In essence, these methods seek stressor values that cause consistent ecological responses within a group (i.e., headwaters) and are different from those observed among different groups. Once these thresholds have been established, they can be used to protect aquatic life uses by relating the thresholds to existing numeric criteria (i.e., pH, dissolved oxygen [DO]), biological assessment outputs, or conceptual models that describe direct and indirect connections to other indicators of designated use support.

Frequency Distribution Methods

FDMs are a common line of evidence used to support NNC (for details and applications see Hawkins et al. 2010, Herlihy et al. 2008, and Paulsen et al. 2008). FDMs use ambient nutrients obtained from waterbodies in a predefined region within which naturally occurring ambient nutrient concentrations are assumed to be as homogeneous as possible—typically ecoregions. FDMs then use predetermined percentiles of the resulting nutrient distributions to establish N and P thresholds. For purposes of defining NNC thresholds, FDMs are frequently restricted to data collected at reference sites because such sites are considered reflective of the natural, or background, ambient nutrient concentrations. Sometimes reference sites are unavailable in a region, in which case NNC thresholds are based on the distribution data from all streams in the region of interest.

The specific percentiles used to establish NNC vary regionally; previous USEPA guidance recommended using the 75th percentile of reference sites or the 25th percentile of all sites. Use of these specific percentiles has been criticized as being overly conservative, because they implicitly assume that nutrient-related problems occur at pre-established rates (e.g., 25% of reference sites would be considered impaired). Perhaps the biggest limitation of FDM methods is that they are not linked to the uses they are intended to protect. Small increases in nutrients do not always cause ecological responses that are deleterious to uses; in some circumstances small increases in nutrients can even be beneficial to some ecological attributes. A departure from the reference condition does not necessarily correspond to degradation of the designated use. Nevertheless, an evaluation of ambient nutrient concentrations remains a useful line of evidence that can be used to support NNC. For instance, FDMs were found to be useful in the derivation of the headwater NNC proposed in this TSD, which are described in Chapter 12.

Mechanistic Models

Mechanistic water quality models, sometimes called process-based models, are another approach used to generate NNC for N and P. These models use predefined mathematical relationships to couple chemical, physical, and biological processes (see Chapra 1997 and Edinger 2002 for reviews). To create these models, the appropriate water quality goals are selected based on parameters most responsive to nutrient enrichment, and then the model is calibrated using monitoring data or literature values to constrain important factors within the model algorithms. NNC can then be "backed out" of the models by asking what concentrations of N or P result in modeled violations to the water quality goals.

Many mechanistic water quality models are used to generate NNC. The Water Environment Federation recently conducted a review and identified 30 potential models that could be used for these purposes (Bierman et al. 2013). These models differ in complexity, and each incorporates and emphasizes the relative importance of different biogeochemical processes.

One advantage of process-based models is that multiple future scenarios can be explored to help support the cause-effect relationships observed in empirical stressor-response relationships (SAB 2010). Another advantage of these models is that they always generate an endpoint—in this context a concentration of N or P that can be used to set NNC. This simplicity can also be a disadvantage because the models vary in their treatment of uncertainty and variability. In general, models, with their focus on processes, most accurately predict physical conditions (i.e., temperature, DO) followed by water chemistry (i.e., nutrients) and then biological endpoints. Perhaps the biggest disadvantage of mechanistic models is that they require site-specific calibration, making them difficult to apply on a regional basis. As a result, DWQ intends to use mechanistic models to develop site-specific NNC or to confirm regional NNC endpoints as TMDLs are developed to address nutrient-related impairments.

Utah's Phased Approach for Numeric Nutrient Criteria Development

DWQ intends to establish appropriate NNC for all surface waters. However, the Nutrient Core Team recommended that NNC be developed and implemented iteratively. DWQ intends to phase the NNC development for different waterbodies over time. The discussion below provides the rationale for initially focusing on headwater streams and discusses how information gleaned from these efforts will be used to inform ongoing nutrient reduction efforts elsewhere.

Prioritization of Headwater Streams for Numeric Nutrient Criteria Development

The majority of Utah's most pristine streams—and many reference sites—are located in headwaters. Therefore, DWQ is initially proposing NNC for headwater streams; details of DWQ's proposed guidelines are presented in Section 2 of this document. Subsequently, DWQ will promulgate NNC for nonheadwater streams on a site-specific basis. Development of programs to address nutrient enrichment in other types of waterbodies such as lakes/reservoirs and wetlands are also ongoing and may include both site-specific and regional NNC. In all cases, the indicators described in this report will become integral to nutrient reduction efforts because they form the basis for nutrient-specific monitoring and assessment approaches. DWQ will use these assessments to identify and prioritize sites for follow-up site-specific investigations.

The decision to prioritize headwaters is based on socioeconomic and technical considerations. Many of Utah's headwater streams are recreation sites or drinking water sources, and they have great economic importance for the state. Utah's antidegradation rules identify areas of the state that contain high-quality headwater streams—often important drinking water sources—and afford them greater protection than other streams in the state (Figure 1.1). In some of these waterbodies (Category 1) the state does not allow any discharges; in the rare cases where discharges are permitted (Category 2), they cannot exceed background conditions of the receiving water. Such waters occur mostly on U.S. Forest Service lands, which have fewer anthropogenic stressors.

DWQ is most confident in the technical application of S-R thresholds headwaters. The numeric

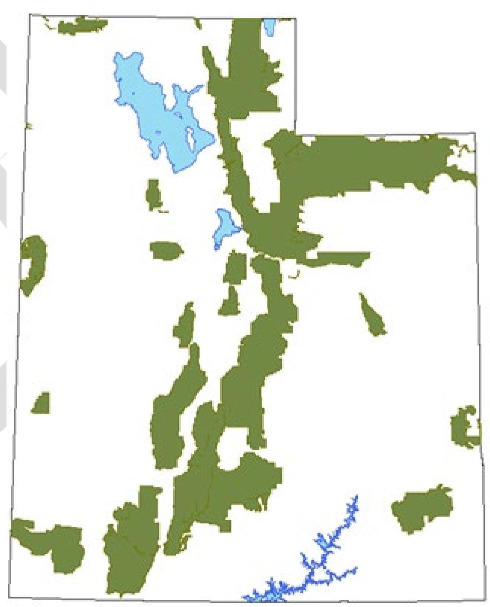


Figure 1.1. This map depicts Utah's Antidegradation Category 1 and 2 boundaries in green. Division of Water Quality is proposing regional nitrogen and phosphorus numeric criteria for these waters (headwaters) prior to developing numeric nutrient criteria for all waters of the state.

thresholds established from these analyses are likely achievable for headwaters, but they may require adjustment for valley streams due to irreversible geomorphic or hydrologic modifications. Moreover, most anthropogenic stressors occur lower in watersheds. Confirmation of indicators downstream of headwaters also requires an evaluation of covariates to confirm that deleterious structural responses are the result of nutrient enrichment. Important covariates to examine include characteristics such as temperature, gradient, substrate size, riparian condition, and other human-caused stressors.

Application of Stressor-Response Measures to Water Quality Assessments

One important outcome of the S-R linkages described in this report is the creation of a suite of tools DWQ can use to identify streams with nutrient-related problems. Using multiple lines of evidence creates a more holistic picture of the varied and often complex responses to cultural eutrophication. Multiple lines of evidence will also provide the data needed to connect increases in nutrients to degradation of uses.

Although different ecological responses to nutrient enrichment could be considered independently, DWQ considers them simultaneously and in the context of defensible conceptual models supported by multiple lines of evidence. In general, this weight of evidence approach also addresses concerns about site-specific covariates that could potentially invalidate regional characteristics. Multiple responses collectively provide insights into linkages among nutrients, processes, and the degradation of designated uses. These evidence-based assessments will also elucidate areas of uncertainty that should be explored to develop effective approaches for reestablishing the biological integrity of degraded streams.

Practically speaking, making assessments using multiple lines of evidence is resource intensive. As a result, DWQ has developed tiered monitoring approaches that use routine monitoring information to screen sites for follow-up investigations. When sites with high levels of N or P are identified, DWQ will conduct additional monitoring to provide detailed measurements of structural and functional responses to eutrophication. If these data identify a nutrient-related impairment, DWQ will use these response indicators to inform the design of site-specific follow-up investigations to establish NNC and appropriate water quality targets that are direct measures of the conditions that need to be restored. Details of DWQ's monitoring methods are provided in Utah's *Strategic Monitoring Plan* (DWQ 2010); detailed assessment methods are described in the methods section of *Utah's Final 2016 Integrated Report* (DWQ 2016). Both documents will be updated to reflect approaches specific to DWQ's nutrient reduction efforts as they are refined and developed.

First Streams, Then Lakes and Reservoirs

DWQ has initially focused on the refinement of N and P indicators and NNC development for stream ecosystems. Lakes and reservoirs are equally important, but in many respects DWQ is already addressing nutrient problems in those waters. Utah's lake assessment methods evaluate several nutrient response metrics—such as the trophic state index, violations of numeric DO criteria, and cyanobacteria dominance—to determine whether nutrients are degrading those waters. When eutrophic lakes are identified, DWQ develops TMDLs that specify in-lake nutrient concentrations and required nutrient load reductions for all sources. Another reason to focus on streams is that effective nutrient management in streams and rivers also affords some protection for lakes and reservoirs because streams are the primary external nutrient inputs to lentic ecosystems. However, in many cases, lakes may actually be more sensitive to excess nutrients than streams, which would require refining upstream NNC through TMDLs. The methods used to address N and P pollution in Utah's lakes and reservoirs need to be refined, and NNC may need to be established for some of these ecosystems, but DWQ has identified development of NNC for headwater streams as the most immediate need.

Great Salt Lake: A Unique Ecosystem That Requires Special Consideration

Specific nutrient requirements pertaining to Great Salt Lake (GSL) must be addressed separately from other lakes because GSL is a unique hypersaline environment in which nutrient requirements and nutrient cycling are very different from freshwater lakes and typical marine environments. Simply put, the deleterious effects of excess nutrients on aquatic life uses that are observed in freshwater lakes may not be applicable to GSL because many sensitive organisms (i.e., fish) do not reside there. Moreover, a nutrient management strategy for freshwater lakes, if applied blindly to GSL, could adversely affect the biota of GSL and the essential ecosystem functions of the lake. Keystone species such as the brine shrimp (*Artemia franciscana)*, identified as a "Species of Protected Aquatic Wildlife" (UAC R657-52-1 and UAC R657-52-11), are fundamentally essential to the ecosystem of GSL. The GSL is a body of water whose ecosystem functions have hemispheric consequences. Because of GSL's unique ecosystem characteristics, hemispheric importance, and protected keystone species, a rigorously detailed and tailored approach to addressing nutrient issues of GSL will be implemented.

Document Organization

This TSD presents the results of several scientific investigations that DWQ conducted (some in collaboration with other individuals or organizations) to support NNC development. The investigations evaluated the relationships between nutrients (stressors) and indicators that quantify important aspects of the condition of aquatic life and recreation uses (responses). The results of these investigations are presented in Section 1. The thresholds derived for these S-R relationships are intended to provide multiple lines of evidence for use in determining what ambient N and P concentrations are most strongly associated with impairments of designated uses. In Section 2 of this TSD, DWQ combines the S-R evidence with several additional lines of evidence to derive and test NNC for headwater streams. Section 3 of this TSD presents the headwater NNC and provides an independent analysis of the NNC. The final section of the TSD (Section 4) provides additional details about the implementation of the headwater NNC and the application of knowledge obtained from the S-R investigations and other investigations to ongoing site-specific NNC development elsewhere. The section includes the results of a sensitivity analysis that DWQ conducted to determine the relative importance of water quality parameters to more accurately calibrate Qual2K models when site-specific standards are developed in the future.

SECTION 1

Relationships between Nutrients and Indicators of Designated Use Support

Chapter 2

EVALUATION OF STRESSOR-RESPONSE RELATIONSHIPS

Key Points

Many different responses to excess nutrient enrichment are potentially deleterious to designated aquatic life uses, recreational uses, and culinary uses.

Some deleterious responses to excess nutrients are direct (e.g., increased abundance of nuisance algae). Others are indirect and occur following nutrient-related alteration of intermediate ecological processes.

Stressor-response models quantify relationships between nutrients (stressors) and measurements of important aspects of designated use conditions (responses).

This section of the technical support document presents the results of several stressor-response analyses that evaluated linkages between nutrients and aquatic life uses (Chapters 3–7) and nutrients and recreational uses (Chapter 8).

The aquatic life use responses evaluated include potential alterations of important ecological processes (functions) and changes to the composition and relative abundance of stream biota.

Recreational uses were evaluated using a survey that showed recipients photographs of streams with varying levels of algal biomass and collected their opinions on the desirability of the streams for recreation activities.

Introduction

This section of the technical support document presents the results of several stressor-response (S-R) evaluations conducted to support numeric nutrient criteria (NNC) development for Utah's headwater streams. This chapter presents a review of some of the broad considerations that informed the specific S-R relationships selected for evaluation and a discussion of the linkages between the responses selected for evaluation and Utah's water quality standards. The chapter includes general caveats and considerations with respect to the interpretation of the resulting S-R relationships. The development and interpretation of S-R thresholds, the challenges inherent to using S-R models derived from survey data to inform NNC, and the approaches employed to address those challenges are also reviewed. This chapter concludes with an overview of the organization of this section, in which the results of the S-R analyses are presented.

S-R models establish quantitative linkages between ambient nutrient concentrations and measures of ecological response that are direct or indirect measures of aquatic life use support (King and Richardson 2004). NNC are then defined based on the nutrient concentration associated with the strongest change in the ecological response. Once these nutrient thresholds have been established, they can be related to protection of aquatic life uses through other existing numeric criteria (i.e., pH, dissolved oxygen [DO]) related to the ecological response that was investigated, biological assessment outputs, or by evaluating the extent to which they are consistent with the predictions of conceptual models that outline known linkages between the nutrients and responses (Table 2.1).

Table 2.1. Links among indicators, nutrients, and Utah's Water Quality Standards (Utah Administrative Code R317-2).

Indicator	Description	Relationship to Nutrients	Tie to Standards
Nutrient-Diffusing Substrates (Chapter 4)	Response of benthic algal production to experimental nigrogen (N) and phosphorus (P) additions.	**Direct**: Algal growth is directly coupled to N and P.	**Narrative**: Excess benthic algae can create "objectionable conditions" (Chapter 8) or conditions detrimental to desirable fish, but only with large accumulation.
Metabolism: Gross Primary Production (Chapter 5)	Reach-scale measure of the total primary production.	**Direct**: Plant and algae growth is directly coupled to N and P.	**Numeric**: Excess production can cause pH to exceed numeric criteria. Also, this metric directly relates to the requirement for total dissolved gasses not to exceed 110% saturation, which is not evaluated in this report. **Narrative**: Can be related to prohibited conditions such as undesirable tastes and odors, surface debris, and undesirable aquatic life.
Metabolism: Ecosystem Respiration (Chapter 5)	Reach-scale measure of the respiration of all plants and animals.	**Direct**: Growth of heterotrophic microbes and fungi is directly coupled to nutrients.	**Numeric**: Directly relates to minimum DO requirements.
Organic Matter Standing Stocks (Chapter 6)	Measures of various standing stocks (storage) of different types of organic matter.	**Both**: Direct if autochthonous (produced within the stream), indirect if allochthonous (from outside the stream).	**Numeric**: Directly relates to minimum DO requirements.
Alteration of the Composition of Macroinvertebrate and Diatom Assemblages (Chapter 7)	Differential responses of sensitive and tolerant taxa to increasing nutrients.	**Indirect**: Assemblage changes occur following nutrient-mediated modifications to physical and chemical habitats.	**Narrative**: Linked to biological assessments, which directly quantify biological integrity, a fundamental Clean Water Act (CWA) goal.

Indicator	Description	Relationship to Nutrients	Tie to Standards
Biological Assessments (O/E) (Chapter 7)	Estimates lost macroinvertebrate diversity—as the ratio taxa observed (O) to those expected (E) without human disturbance.	**Indirect**: Assemblage changes occur following nutrient-mediated modifications to physical and chemical habitats.	**Numeric**: Designated use descriptions require protection of the food web. **Narrative**: References to undesirable or nuisance organisms.
Recreational Use: Filamentous Algae (Chapter 8)	Survey of Utahns to evaluate the extent to which large algal blooms make river recreation less desirable.	**Direct**: Algae growth is directly coupled to N and P. Filamentous algae become dominant in high-nutrient, cobble-bedded streams.	**Numeric** (proposed): Proposal to protect recreational uses with benthic algal biomass (chlorophyll-*a* [chl-*a*] or ash free dry mass [AFDM]) and filamentous algae cover. **Narrative**: References to undesirable or nuisance organisms.

Selection of Candidate Ecological Responses

Cultural Eutrophication and Aquatic Life Uses

General Effects of Cultural Eutrophication on Stream Ecosystems

Anthropogenic eutrophication can have numerous deleterious impacts to aquatic life, recreation, or drinking water uses. Several authors have reviewed the multiple pathways that link excess N and P to degradation of aquatic ecosystems (e.g., Dodds 2006, Smith et al. 1999, USEPA 2010). These links are illustrated in the conceptual model in Figure 2.1. While not appropriate for all streams or responses, conceptual models like these are useful because they include several important considerations relevant to the derivation of NNC, including the importance of numerous sources, modifying factors (i.e., covariates), ecosystem responses, biological responses, and interacting stressors. While a literature review of nutrient effects on ecosystems is unnecessary here, a summary of several investigations that most directly relate to the state's approaches for development of NNC and associated assessment tools is provided in this chapter.

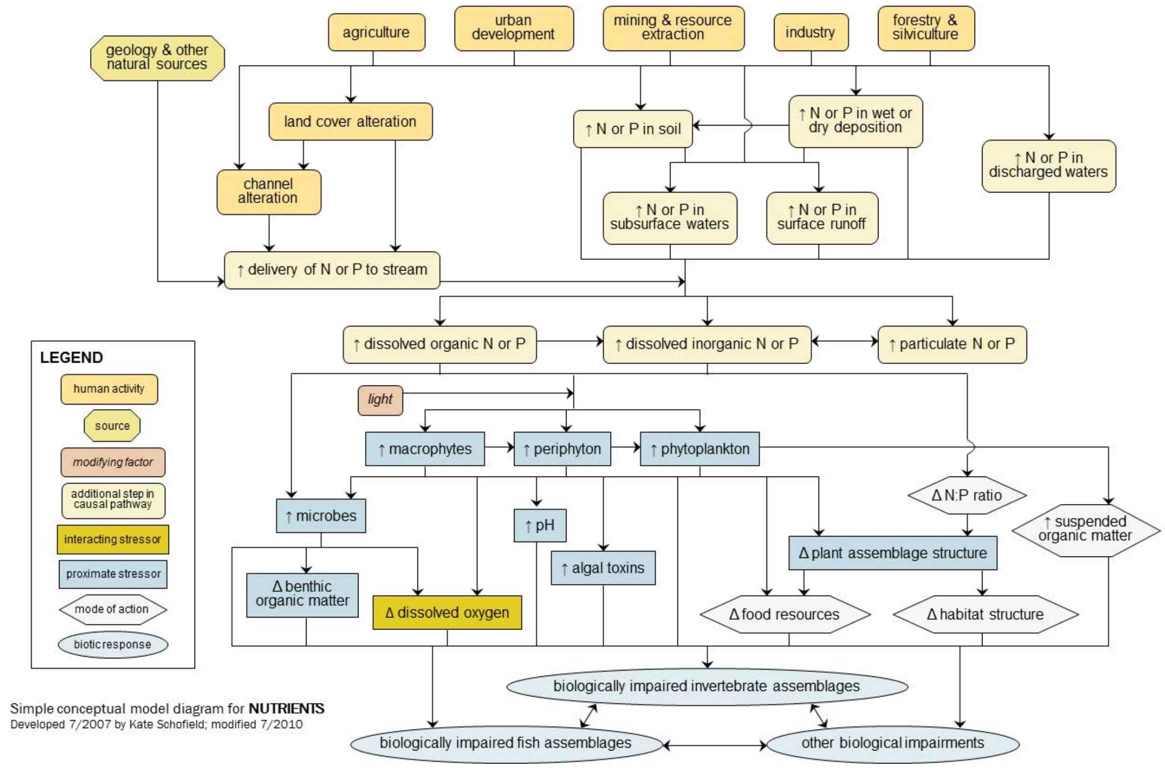


Figure 2.1. Conceptual linking of nutrient sources to ecological modifications that have the potential to cause degradation of aquatic life uses (USEPA 2000).

One principal pathway through which excess nutrients can degrade stream ecosystems is via increases in primary production leading to increased growth of macrophytes or algae. In larger, soft-bedded rivers this growth is typically manifest either as macrophytes or phytoplankton (Vanote et al. 1980). In contrast, in cobble-bedded streams, a shift from diatoms to filamentous algae is often observed (Slavik et al. 2004). Consequently, investigators frequently use alterations of algal assemblage composition to document nutrient-related effects (see Chapter 7). Initially—as nutrients increase beyond background concentrations—an increase in primary production poses few problems and can even be beneficial to stream ecosystems. Eventually, however, excess growth can cause several problems, the severity of which depends greatly on the site-specific biological characteristics (Biggs et al. 2000, Haapala et al. 2001) and physicochemical characteristics (Biggs 2000, Lohman et al. 1992, Townsend et al. 2012).

One important ecological effect of eutrophication is reduced DO concentrations within streams. Increases in primary production correspondingly increase photosynthesis rates and daytime DO production. At night, all organisms continue to respire, which consumes DO. At night, since this consumption is not offset by the DO produced by photosynthesis, water column DO concentration declines. If atmospheric reaeration cannot compensate for these DO losses, then DO concentration within the stream can fall to levels that are unhealthy, even deadly, to some stream biota. Eutrophication, caused by excess N and P, exacerbates low nighttime DO because heterotrophic bacterial and fungal productivity can also be stimulated by nutrient enrichment (Chapter 5). Also, increases in plant or algae growth that sometimes occur in high-nutrient environments creates additional carbon, which eventually leads to greater secondary production and an associated increase in respiration (Chapters 5 and 6). Particularly in temperate streams, these carbon increases have the potential to create low DO conditions during autumn senescence of algae or macrophytes, which provides a large pulse of labile carbon (Suplee et al. 2012). Overall, the close coupling of DO to both production and respiration provides a relatively simple way to obtain reach-scale estimates of both processes (Chapter 5).

Direct acute or chronic effects of low DO to stream biota are not the only deleterious impacts to stream ecosystems. Low DO decreases the sequestration rates of some heavy metals to stream sediment; this creates additional threats to stream biota, especially considering that they are already under stress. Similarly, low DO sometimes interacts with other stressors (i.e., ammonia), which results in an increased threat to stream biota.

Excessive primary production (Chapter 5) degrades aquatic life directly through alterations to stream food webs and indirectly through degradation of stream habitat. In cobble-bedded streams, increases in primary production is manifest as high periphyton biomass, often resulting in a shift in the base of the food web from one dominated by diatoms to one dominated by filamentous algae. Special physiological adaptations are required to consume filamentous algae, or the organic matter trapped within, which alters the abundance and distribution of macroinvertebrate taxa (Dudley et al. 1986). Fish can benefit from the cover that filamentous algae provide, but negative effects are associated with habitat degradation. A high abundance of filamentous algae alters benthic flow characteristics and traps fine sediment (Slavik et al. 2004). Eventually this fine sediment fills benthic interstitial spaces, which is critical habitat for macroinvertebrates that are the base energy source of stream fishes (Wallace and Webster 1996). These changes also alter the types and relative abundance of organic matter stored in the stream, which is another potentially useful enrichment response (Chapter 6). People are also affected by excessive periphyton growth, because it can physically interfere with recreational activities, such as fishing, and cause streams to be less aesthetically appealing for recreational uses (Chapter 8; Suplee et al. 2012).

Nutrient enrichment can also degrade aquatic life uses in low-gradient, soft-bedded streams, although the specific changes differ from those observed in cobble-bedded streams. At relatively low levels of enrichment, the water column remains relatively clear, and primary production is frequently dominated by macrophytes; these conditions create high quality habitat for stream biota (Fritz et al. 2004). As nutrient enrichment continues to increase, increases in phytoplankton abundance can occur, shifting the base of the food web from benthic to pelagic sources (Chapter 6; Vanote et al. 1980). This shift in the food web base decreases water clarity and can lead to further declines in macrophyte abundance. Higher organisms are also affected by these changes, and the composition of macroinvertebrates and fish shifts to species better adapted to a pelagic-based food web (Chapter 7).

Eventually, incremental increases in nutrients cannot further increase primary production because other factors become more important in limiting growth of plants and algae; factors such as light (channel shading) or residence time (slope) can constrain growth (Bernot and Dodds 2005). Similarly, an increase in either N or P can potentially switch the nutrient that most limits production (Chapter 4; Dodds et al. 1997); under natural conditions, saturation of ***both*** N and P rarely occurs. Thus, nutrient saturation concentrations can be an indication of human-caused nutrient enrichment, especially in circumstances where saturation of both macronutrients is observed. The saturation concentrations of N and P also have important restoration considerations because future nutrient reductions would be unlikely to improve stream conditions until they result in ambient concentrations below saturation thresholds.

Nutrients can be transported downstream, which sometimes causes problems far from nutrient sources. As a result, nutrient reduction efforts must also consider factors related to the transport and storage of nutrients. Once nutrients enter a stream, they can be retained and accumulate locally in stream sediment, groundwater, or in the tissue of stream biota. However, much of this storage is transitory, and many of these nutrients are ultimately transported downstream. The importance of hydrologic transport of nutrients to downstream sources is described by the nutrient spiraling conceptual model (Newbold 1981). In this model, an N or P molecule enters a stream and is eventually taken up by stream biota (e.g., biological uptake by periphyton, chemical adhesion to stream sediment). Eventually, the nutrient molecule is released in dissolved form to the water column where is transported some distance downstream before being taken up again, completing one nutrient spiral. The uptake rates and lengths of nutrient spirals have been used as metrics for understanding the extent to which nutrient enrichment has altered biogeochemical cycles. Alterations to these cycles differ appreciably among streams. Many of these differences result from natural changes in stream conditions, but human activities are also important. For instance, net uptake length in highly enriched streams can be an order of magnitude greater than less degraded systems (Haggard et al. 2005). Such patterns may not be universal, but they do highlight the importance of nutrient transport to aquatic resource management because addressing nutrient enrichment in these streams would prevent local problems while more effectively protecting ecosystems downstream.

Selection of Candidate Responses

Ecological responses to nutrient enrichment can be categorized as reflecting the effects of nutrients on stream ecosystem structure or function. Structural ecological responses quantify the effects of nutrients on the composition of stream biological assemblages. In contrast, functional responses measure stocks of materials and the processes involved in moving energy among trophic levels—whether through or mediated through the stocks of materials. As with all attempts at ecological categorizations, the distinction between structural and functional processes is not black and white. It is possible that alterations to macroinvertebrates, for instance, could disproportionally affect one functional feeding group over another, which could be characterized as either a structural or functional change to the ecosystem. The distinction is useful as a reminder that maintenance of biological integrity involves considering the species resident in the stream and those processes involved in the maintenance of healthy ecosystems.

Division of Water Quality (DWQ) elected to consider both types of effects when evaluating its NNC S-R efforts.

Based on available literature, states have mostly relied on structural ecological responses for purposes of NNC development (Florida Department of Environmental Protection 2012, Maine Department of Environmental Protection 2009, Minnesota Pollution Control Agency 2008, Weigel and Robertson 2008). Several aquatic assemblages that are known to be sensitive to nutrient enrichment have been evaluated in various investigations; such assemblages include macroinvertebrates (i.e., King and Richardson 2003), algae (i.e., Dodds et al. 2002), and (less commonly) fish (i.e., Wang et al. 2007). The focus on structural responses for NNC development reflects the long-standing use of these indicators in biological assessment programs by both state and federal agencies, including DWQ (see DWQ 2010, 2014). USEPA guidance on S-R models states that the ideal response would meet the dual goals of measuring whether or not a designated use is supported and being relatively sensitive to increases in N or P. Many states have developed biological assessment programs using structural indicators of condition because they are integrative through time and across space and are, therefore, often reflective of broad stream conditions (Barbour et al. 2000, Karr 1981), which is the greatest strength of these responses in informing NNC.

Despite their strength in biological assessment programs, measures of structural condition—particularly those based on macroinvertebrates and fish—may not be the best way to quantify ecological responses to anthropogenic eutrophication. Except under extreme conditions (i.e., ammonia toxicity), nutrients are usually not directly toxic to aquatic life; instead, intermediary effects of increased nutrients on stream processes—or functions— subsequently cause alterations to the structure of stream biota. Responses that quantify these intermediate responses are direct, and potentially more accurate and sensitive, responses to anthropogenic nutrient enrichment. Indeed, several researchers have proposed augmenting biological assessments with direct measures of ecosystem processes (Grace and Imberger 2006). However, states, including the State of Utah and DWQ, have not traditionally incorporated such measures into routine stream monitoring programs due to resource constraints.

DWQ has elected to evaluate a variety of functional and structural response thresholds to identify those that are sensitive to nutrient enrichment and indicate changes that may affect the designated use (Table 2.2). Previous studies have demonstrated that each of these responses is potentially sensitive to nutrient enrichment. Each response is also related, directly or indirectly, to elements of Utah’s water quality standards (Table 2.1).

Table 2.2. Summary of indicators used to develop nutrient thresholds in the proposed headwater numeric nutrient criteria.

Functional Indicators	
Nutrient Saturation	These thresholds, derived from nutrient-diffusing substrates, quantify the concentration of total nitrogen (TN) and total phosphorus (TP) where, on average, additional nutrients did not cause an increase in algal growth. At these thresholds, other factors, such as light, substrate, or CO_2, cause additional algal growth.
Stream Metabolism: Gross Primary Production (GPP)	GPP measures the total amount of oxygen produced by photosynthesizing plants and algae each day (g O_2/m^2/day). Nutrient thresholds derived from GPP are the concentrations that were associated with streams with relatively low, moderate, and high rates of GPP. DWQ proposed to use this as a water quality criterion paired to nutrient criteria.
Stream Metabolism: Ecosystem Respiration (ER)	ER measures the oxygen consumed either through the processing (oxidation) of organic matter to CO_2 or by plant and algae growth (g O_2/m^2/day). Many stream organisms—including bacteria, protozoa, and fungi—rely on organic carbon as an energy source for cellular metabolism and growth, and these processes consume oxygen. DWQ proposes to use this metric as a water quality criterion paired with nutrient criteria.
Organic Matter Standing Stock	Organic matter standing stocks quantify, as g C/m^2 or AFDM/m^2, the amount of organic matter, excluding larger particle and macrophytes, in stream reaches. This measure provides an estimate of the amount of material available to feed bacterial, protozoan, and fungal respiration. Nutrient thresholds were derived as the concentrations that, on average, distinguished among streams with relatively low, moderate, and high organic matter standing stocks.
Structural Indicators	
TITAN (nCPA): Diatoms	Threshold indicator taxa analysis (TITAN; Baker and King 2010) is a method that calculates indicator scores that capture the occurrence, abundance, and directionality of species responses to stressors. As proposed for use here, the method (nonparametric change point analysis [nCPA]) then uses these indicator scores to determine the nutrient concentration that describes the point at which statistically significant changes in species abundance occur. For the proposed headwater NNC, TITAN scores capture biological changes associated with diatom taxa—a diverse assemblage of algae that are known to be sensitive to nutrients.
Macroinvertebrate Biological Assessments: O/E	DWQ currently uses macroinvertebrate-based river invertebrate prediction and classification system (RIVPACS) models to evaluate biological integrity and to determine whether streams are meeting their designated aquatic life uses. The output of the models, O/E, is a ratio of the number of macroinvertebrates that were actually observed at a site compared with the number of species that were predicted to occur in the absence of human-caused stressors. Nutrient thresholds were derived for concentrations that best distinguished between streams in degraded and nondegraded conditions.

Interpretation of Stressor-Response Models: Caveats and Considerations

The extent to which excessive nutrients result in any alteration of any ecological response depends greatly on site-specific characteristics. Variations in natural conditions (e.g., channel shading, water temperature) cause among-stream differences in responses to nutrient enrichment. For instance, higher-gradient streams are buffered against potential DO problems because they typically have higher canopy cover, which lowers primary production and increases atmospheric reaeration. Higher-gradient streams are also less likely to accumulate N, P, or C, which affords them natural protection from chronic nutrient inputs. Structural responses are also affected by these natural environmental gradients. For example, among-stream differences in temperature or slope directly alter the distribution of macroinvertebrate or algae taxa.

Human activities may cause changes to the specific conditions at a site; those changes may then modify the magnitude of responses to nutrient enrichment. For instance, people sometimes remove riparian vegetation, which can increase primary production rates. Other landscape-level changes, particularly urbanization, alter the transport and storage of nutrients in streams (Paul and Myer 2001). These land uses also cause other sources of stress that can have both synergistic and antagonistic effects to stream structure and function. When feasible, nutrient restoration plans should consider and incorporate improvements to all the degraded physicochemical characteristics that operate in concert with nutrients to degrade stream conditions. However, in some cases such changes are practically irreversible, in which case new goals—water column concentrations or ecological response thresholds—will need to be established.

The study investigated two classes of ecological responses to evaluate S-R relationships for aquatic life uses.

Functional Indicators: Measures of processes (or properties) that focus on physical-biological-ecological linkages, in particular those that describe the flow of energy through ecosystems.

Structural Indicators: Measures of the composition and relative abundance of individuals in aquatic assemblages, such as macroinvertebrates, diatoms, and fish.

Section Organization

This section of the report summarizes studies conducted by DWQ to establish linkages between nutrients and designated uses. The first part of the section focuses on relationships between ambient stream nutrient concentrations and indicators of aquatic life use support. Aquatic life use evaluations include a DWQ investigation establishing relationships between nutrients and stream ecosystem functions (Chapters 3–6). Chapter 3 provides an overview of the study design and a summary of the analytical methods that are common to the studies in Chapter 4–6. Functional responses are assessed in three chapters: nutrient limitation and saturation in Chapter 4, whole stream metabolism in Chapter 5, and organic matter standing stocks in Chapter 6. Chapter 7 presents the results of an analysis using existing biological assessment data to evaluate how nutrients alter the composition of diatoms and macroinvertebrates. Finally, Chapter 8 presents the results of a survey evaluating the potential impacts of excessive algae growth on recreational uses.

Chapter 3

DEVELOPMENT AND TESTING OF FUNCTIONAL ECOLOGICAL RESPONSES AS INDICATORS OF NUTRIENT ENRICHMENT

Key Points

Measures of ecosystem function are promising indicators of nutrient enrichment but are not routinely collected by Division of Water Quality as water quality indicators.

Division of Water Quality conducted a special study at 35 streams (the reference sites) to obtain data on ambient nutrient concentrations and several measures of ecosystem function: nutrient limitation and saturation, whole stream metabolism, and organic matter standing stocks.

Streams in this study vary greatly in their relative extent of nutrient enrichment. Total nitrogen ranges from 0.11 to 14.72 mg/L, and TP ranges from 0.003 to 7.89 mg/L.

Introduction

The next four chapters of this report provide the results of a special investigation that the Division of Water Quality (DWQ) conducted to evaluate whether several measures of stream ecosystem functions could quantify the effects of nutrient enrichment on stream ecosystems. Specifically, the study evaluated three functional responses of streams with varying nitrogen (N) and phosphorus (P) concentrations: nutrient limitation and saturation, stream metabolism, and organic matter standing stocks. Initially, the principal study objective was to support development of statewide numeric nutrient criteria (NNC). As the study progressed, it became clear that for many streams NNC would need to be derived on a site-specific basis because traditional classification methods did not sufficiently minimize natural variation in ecosystem responses at a regional scale. Under circumstances where direct application of regional thresholds to specific sites is not appropriate, DWQ will use the field and analytical methods developed for this study to augment ongoing monitoring and assessment programs to more accurately identify streams with nutrient-related water quality problems.

This chapter describes the general study design and the methods that apply to all three indicators, including collection of the underlying water chemistry data. Methods specific to each indicator are presented in the relevant chapter and in the standard operating procedures (SOPs) in the report appendices.

Methods

Site Selection

Functional responses (stressor-response [S-R] data) were evaluated from 35 wadeable stream sites in central and northern Utah. The streams selected for analysis represent the known variation in N and P values observed in Utah streams. In all, 17 reference sites were selected, along with 9 sites immediately above and 9 sites immediately below the mixing zone of publicly owned treatment works (POTW) effluent discharges (Figure 3.1). Although the study methods proposed measuring all responses at all sites, in some cases field logistics precluded measurement of some indicators. Thus, the number of sites evaluated differs slightly for different responses.

Upstream locations were expected to represent streams predominantly enriched by nonpoint

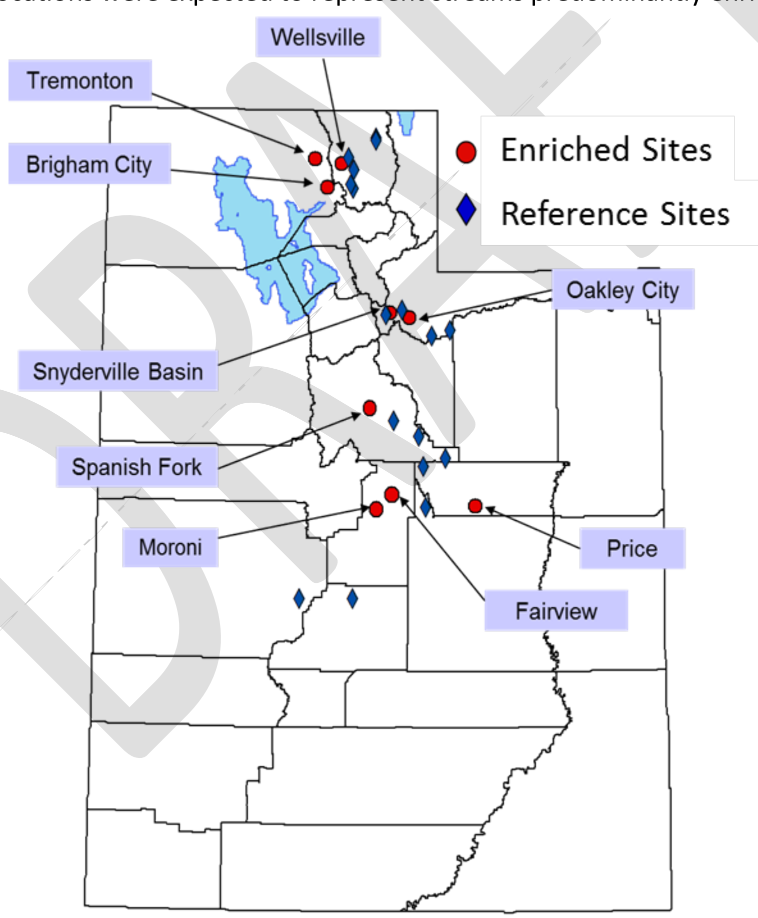


Figure 3.1. Site locations for the 2010 ecological impacts of nutrients study. Enriched sites are a combination of two study reaches above and below a publicly owned treatment works discharge (labeled). Reference sites represent minimally enriched stream locations.

source nutrient inputs, whereas downstream locations were expected to also include point source inputs. The sampling design was not developed to target POTW sources. DWQ recognizes that these wastewater treatment plant (WWTP) reclamation facilities are not the only, nor always the largest, sources of stream nutrients. Whenever possible, existing data were used to select reference sites with geomorphic characteristics (i.e., slope, width, substrate size) similar to the sites upstream and downstream of the POTWs (Table 3.1).

Table 3.1 Streams sampled in the functional assessment study.

Waterbody	Location	Code	STORET	Type
Blacksmith Fork	At U101 Crossing	BLACKFK	4905440	Reference
Box Elder Creek+	Above Brigham City WWTP	BEC-AB	4901180 B	Above
Box Elder Creek+, *	Below Brigham City WWTP	BEC-BL	4901180 D	Below
Diamond Fork	Below Palmyra Campground	DIAFK	4995665	Reference
Dry Creek	Above Spanish Fork WWTP	DCSP-AB	4996020 B	Above
Dry Creek	Below Spanish Fork WWTP	DCSP-BL	4996020 D	Below
Fish Creek	Above Scofield	FISHCK	5931650	Reference
Huntington Creek+	0.5 Miles below Guard Station	HUNTCK	4931230	Reference
Little Bear River	West of Avon at Road Crossing	LBRAVON	4905700	Reference
Little Bear River	Above Wellsville Lagoons	LBRW-AB	4905600 B	Above
Little Bear River	Below Wellsville Lagoons	LBRW-BL	4905600 D	Below
Logan River	At 1000 West	LOGR1000	4905140	Reference
Logan River*	By the Dugway	LOGRDUG	4905260	Reference
Logan River	Below Twin Bridges	LOGRTB	4905195	Reference
Malad River*	Above Tremonton WWTP	MRTRE-AB	4902710 D	Above
Malad River*	Below Tremonton WWTP	MRTRE-BL	4902710 D	Below
McLeod Creek	At Swaner Nature Preserve	KIMBALL	4925442	Reference
North Fork Chalk Creek	Above South Fork	NFCHLK	4940201	Reference
Price River+, *	Above Price WWTP	PRP-AB	4932370 B	Above
Price River+, *	Below Price WWTP	PRP-BL	4932370 D	Below
Price River	Below Kyune A Railroad Tunnel	PRICER	4932815	Reference
South Fork Lower Bear River	Above East Fork	SFKLBR	4905740	Reference
Salt Creek+, *	Below Salt Canyon	SALTCK	4995355	Reference
San Pitch River	Above Fairview City WWTP	SPRFV-AB	4946830 B	Above
San Pitch River	Below Fairview City WWTP	SPRFV-BL	4946830 D	Below
San Pitch River	Above Moroni Feed WWTP	SPRM-AB	4946970 B	Above

Waterbody	Location	Code	STORET	Type
San Pitch River	Below Moroni Feed WWTP	SPRM-BL	4946970 D	Below
Silver Creek	Above Synderville-Silver Creek WWTP	SCSNYD-AB	4926790 B	Above
Silver Creek	Below Synderville-Silver Creek WWTP	SCSNYD-BL	4926790 D	Below
Tie Fork	2 miles above Hwy 6	TIEFK	4995928	Reference
Unknown Stream	Above Provo Falls	UKMURD	4999050	Reference
Upper Provo River	At North Fork	UPRNFK	4998700	Reference
Weber River	Above Oakley City WWTP	WROAK-AB	4928010 B	Above
Weber River	Below Oakley City WWTP	WROAK-BL	4928010 D	Below
Weber River	Above Rockport	WEBR	4927250	Reference

Notes: Waterbody and location describe the catchment and specific locations within the catchment, respectively. Code and STORET (STOrage and RETrieval water quality database maintained by USEPA) are abbreviated site designations and are included here for later reference. The type column distinguishes reference sites from those located upstream (above) and downstream (below) from the publicly owned treatment works. Footnotes: + = no NDS response data, * = no metabolism response data.

S-R relationships were assumed to be more strongly associated with ambient nutrient concentrations than with different nutrient sources. Upstream-downstream comparisons were not conducted for several reasons: (1) The primary objective was to derive general regional thresholds, not site-specific relationships. (2) In some cases it was not clear whether direct upstream-downstream comparisons were appropriate due to locally important covariates. For instance, stream channel characteristics (i.e., width, slope) sometimes differed between upstream and downstream locations. (3) In all cases, but to varying degrees, hydrologic characteristics differed due to the influence of the discharge. Eventually, site-specific upstream-downstream comparisons will help inform site-specific investigations, but additional evaluations will be required to determine the extent to which physical and hydrologic factors alter observed functional responses.

Water Chemistry

In this report, covariate refers to any secondary variable that affects—either positively or negatively—the relationship between nutrients and ecological responses. Sometimes covariates result from natural environmental gradients (e.g., channel shading, stream gradient). In other circumstances, covariates may arise from stressors, other than nutrients, that cause similar ecological responses to occur. The former can be accounted for in S-R relationships, whereas the latter complicate cause-effect conclusions (see Chapter 15).

Individual study sites representative of stream reaches were identified (Figure 3.2). At each reach DWQ monitoring staff collected water chemistry grab samples at the top and bottom of the stream reaches. Depending on the potential for other nutrient inputs (i.e., irrigation return ditches, inflows), additional collection stations were established downstream of each POTW largely for the purpose of constructing water quality models (which are presented in Chapter 16 of this technical support document). Water chemistry samples were collected whenever field crews were on site, which occurred at least three times at each site during the summer of 2010. Field crews followed DWQ SOPs to collect separate samples for total and filtered nutrient analyses. Samples for dissolved analytes were field filtered through 0.45 µm membrane filters (Millipore Corporation). Samples were preserved by freezing them immediately following collection.

The Utah State Biogeochemical Laboratory thawed samples and immediately conducted nutrient

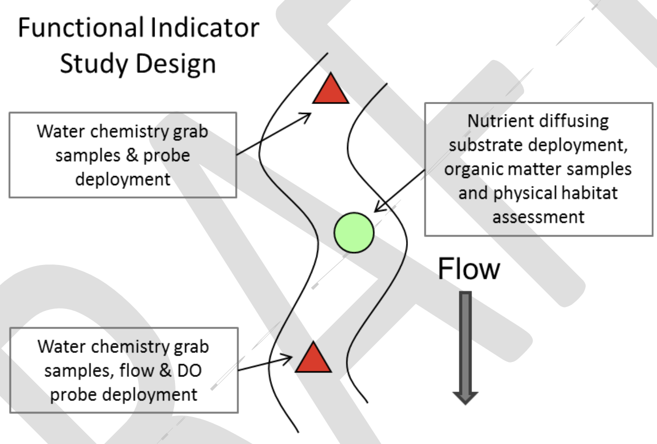


Figure 3.2. Study design for the 2010 ecological impacts of nutrients study to develop functional indicators of nutrient pollution. Sampling block design was repeated at all site locations.

chemical analyses for total N (TN), total dissolved N, total P (TP), and total dissolved P with persulfate oxidation followed by standard colorimetric analysis (Valderrama 1981). The lab also processed these samples with standard colorimetric analysis to obtain nitrate+nitrite (USEPA method 353.4), ammonium (USEPA method 349), and phosphate (USEPA method 365.5). Unless otherwise noted, nutrient data throughout the report are the summertime average of a minimum of six spatial and temporal composites for each reach.

Other water chemistry constituents, including dissolved metals and major anion/cations, were obtained from the Utah State Health Laboratory. Some of these data were used to evaluate the influence of other factors on S-R relationships, but most of these data will be used in follow-up site-specific investigations to evaluate the potential influence of other potential stressors.

Physical and Hydrologic Characteristics

Each site was surveyed using Utah's Comprehensive Assessment of Stream Ecosystem protocols (USEPA 2007), which includes quantitative and qualitative measures of riparian and in-stream physical and hydrological characteristics. Standard DWQ procedures were followed to obtain discharge measurements concurrent with water chemistry samples. Discharge was calculated using the velocity-area method. Velocity was measured with FlowTracker Handheld Acoustic Doppler Velocity meters (SonTek®). Physical habitat data consisted of numerous measures of channel morphology, including wetted and bankfull widths, gradient, substrate size, and bank angles. Channel shading at the banks and center of the channel was measured at 11 equidistant transects along a stream length of 40 times the wetted width of each stream.

The hydrologic and physical data were primarily used to evaluate the extent to which the streams in this investigation were generally representative of Utah streams and to evaluate the potential influence of covariates on stream S-R relationships. Importantly, covariate analyses were broad and only accounted for their general influence on indicators at all study sites. In practice, the influence of these covariates is site- specific. In other words, these analyses indicate whether something like slope generally influences a response metric, but it does not indicate the extent of this influence for a specific site.

Deployment of Multi-Parameter Sondes

Field crews deployed water quality probes (YSI 6600V2 or 600 OMS V2) for a minimum of 48 hours at each upstream and downstream station during the peak growing season (July through early September). All sondes recorded measures of pH, temperature, dissolved oxygen (DO), and specific conductance every 5 minutes. The larger sondes (YSI 6600V2) also recorded chlorophyll and turbidity concentrations. All parameters were calibrated simultaneously on site and immediately prior to deployment. Independent measures, from a recently calibrated sonde, were collected at the end of each deployment to evaluate whether sensor fowling resulted in parameter drift.

Photo 3.1. Example of a water quality probe.

To improve the accuracy of calculations, two stream metabolism sondes were placed at the top and bottom of each reach with optimal between-sonde spacing—the approximate distance required for half of the DO molecules to exchange with the atmosphere. Estimates of optimal sonde spacing were derived from the following relationship (Grace and Imberger 2006):

$$D = \frac{v^{0.33}}{50.8 \times (d^{-0.85})}$$

where:

D = probe distance (km)
v = stream velocity (cm/sec)
d = mean water depth (cm)

Nutrient-Diffusing Substrates

Nutrient-diffusing substrates (NDSs) were deployed in a glide located approximately midway within each reach. NDSs were positioned within each glide at a location that maximized exposure to sunlight (typically the center of the channel). Additional details on collection methods and results are provided in Chapter 4.

Organic Matter Standing Stocks

Volumetric samples of various stores of organic matter standing stock were collected along a minimum 50 m stream segment approximately midway within each reach. These samples were collected at a separate event at the end of the field season. Additional details on collection methods and results are provided in Chapter 6.

Discussion

Mean nutrient concentrations among study sites varied considerably for both TN (0.11–14.72 mg/L) and TP (0.003–7.89 mg/L) (Table 3.2). For context, these TN and TP concentrations were compared with statewide estimates; the nutrient concentrations at the study sites span the range of predicted statewide observations (Figure 3.3). The sample set intentionally included a disproportionate number of sites with unusually high and low (reference site) nutrient concentrations because these sites are relatively rare; this approach avoided giving a single site too much leverage in regional S-R relationships. Selecting representative populations of streams with relatively low, moderate, and high nutrient concentrations provides better estimates of the range of functional responses present in Utah streams. Streams in the sample also differed with respect to other characteristics that have the potential to confound S-R relationships, and the effects of these covariates on functional responses were also evaluated. The inclusion of streams with diverse characteristics, located throughout the state, helps ensure that the S-R models are broadly applicable. Missing important covariates would generally weaken, not strengthen, regional S-R models due to systematic bias (USEPA 2010). In other words, unaccounted-for covariates can cause a systematic bias that obscures linkages between nutrients and responses and hinders the ability of S-R models to demonstrate that nutrients are the principal cause of any deleterious responses observed.

Table 3.2. Nutrient concentrations and physical characteristics of all upstream and downstream sites and representative reference sites that were sampled for the functional response study.

Watershed	Location	TN (mg/L)		TP (mg/L)		Area (mi^2)	Forested (%)	Slope (%)	Channel Shading (%)	Mean Depth (cm)	Mean Width (m)	Discharge (cfs)
		Mean	SD	Mean	SD							
Box Elder Creek	Above Brigham City POTW	0.404	0.117	0.058	0.038	37.9	10.5	0.3	25.0	11.1	2.99	0.3
Box Elder Creek	Below Brigham City POTW	6.351	4.346	0.838	0.716	38.1	10.5	0.1	92.1	30.5	5.17	2.7
Blacksmiths Fork	At HWY 101	0.188	0.040	0.008	0.001	269.0	31.1	1.3	77.3	40.2	8.26	116.1
Dry Creek	Above Spanish Fork POTW	2.216	0.895	0.135	0.063	18.1	35.9	0.3	14.8	45.6	4.43	2.7
Dry Creek	Below Spanish Fork POTW	11.286	3.083	1.894	0.470	19.2	34.2	0.2	0.0	72.0	5.75	4.2
Diamond Fork	At Palmyra Campground	0.410	0.063	0.084	0.041	139.0	57.4	1.7	5.6	24.2	11.69	68.4
Fish Creek	Above Scofield Resevoir	0.246	0.049	0.063	0.072	63.2	67.4	0.3	11.0	14.9	9.98	5.9
Huntington Creek	In Huntington Canyon	0.455	0.063	0.015	0.003	46.0	69.8	1.5	16.6	20.9	6.09	18.3
Kimball Creek	At Swaner Nature Preserve	0.279	0.057	0.028	0.004	29.9	68.7	0.3	8.6	25.2	4.03	15.2
Little Bear River	West of Avon	0.343	0.234	0.021	0.009	77.6	21.7	0.8	34.1	16.7	7.86	3.5
Little Bear River	Above Wellsville Lagoons	1.175	0.183	0.075	0.014	250.0	16.7	0.2	96.6	21.7	7.85	9.6
Little Bear River	Below Wellsville Lagoons	1.085	0.163	0.084	0.010	255.0	16.5	0.2	5.2	60.4	8.62	27.9
Logan River	At 1000 W	0.424	0.080	0.023	0.005	255.0	33.6	0.3	69.7	41.8	11.27	48.0
Logan River	Below the Dugway	0.139	0.030	0.012	0.002	135.0	40.3	1.9	16.0	27.4	11.99	116.1
Logan River	Below Twin Bridges	0.132	0.025	0.012	0.004	141.0	40.1	1.3	19.7	29.3	10.63	73.0
Malad River	Above Tremonton POTW	2.828	0.570	0.236	0.051	689.0	5.2	0.1	41.1	57.4	8.45	26.0
Malad River	Below Tremonton POTW	3.897	1.204	0.445	0.144	691.0	5.2	0.1	37.0	65.8	9.07	27.0
North Fork of Chalk Creek	Above South Fork confluence	0.161	0.040	0.006	0.003	27.5	81.6	3.4	80.0	11.3	5.1	13.5
Price River	Below Kyune	0.388	0.048	0.050	0.014	327.0	61.6	0.9	4.0	33.6	12.65	103.1
Price River	Above Price POTW	0.710	0.136	0.289	0.368	720.0	44.6	0.3	11.6	34.5	6.26	21.3
Price River	Below Price POTW	2.950	1.097	0.732	0.293	849.0	42.1	0.2	12.3	32.5	7.27	19.0
Salt Creek	Above mouth of Salt Canyon	0.154	0.112	0.016	0.004	95.1	48.3	2.5	71.0	20.6	2.55	3.9
Silver Creek	Above Snyderville Basin POTW	0.319	0.121	0.015	0.008	15.9	40.2	0.2	10.7	47.2	1.60	1.1
Silver Creek	Below Snyderville Basin POTW	14.717	4.613	2.212	0.715	17.2	38.9	0.5	1.5	45.0	2.41	2.0

Watershed	Location	TN (mg/L)		TP (mg/L)		Area (mi^2)	Forested (%)	Slope (%)	Channel Shading (%)	Mean Depth (cm)	Mean Width (m)	Discharge (cfs)
		Mean	SD	Mean	SD							
South Fork of Little Bear River	Below Davenport Creek	0.229	0.040	0.017	0.003	63.0	20.9	1.6	25.4	23.3	8.19	15.8
Spanish Fork River	Above Fairview POTW	1.324	0.215	0.019	0.009	96.5	56.5	0.6	4.4	20.2	5.08	5.5
Spanish Fork River	Below Fairview POTW	1.677	0.232	0.078	0.035	98.0	55.9	0.5	10.3	32.1	4.25	6.7
Spanish Fork River	Above Moroni POTW	1.232	0.116	0.033	0.008	227.0	42.9	0.3	6.9	31.4	5.77	0.8
Spanish Fork River	Below Moroni POTW	10.416	1.194	7.897	1.449	228.0	42.6	0.2	20.7	41.6	6.40	2.0
Tie Fork	Two miles above HWY 6	0.118	0.011	0.007	0.002	16.8	64.6	2.4	24.6	15.4	2.30	2.5
Unnamed Stream	In Murdock Basin	0.109	0.017	0.004	0.001	3.7	66.5	4.7	48.9	14.6	3.00	0.8
Provo River	At North Fork Trailhead	0.113	0.023	0.003	0.003	66.6	77.8	7.1	4.3	24.5	13.94	54.4
Weber River	Above Rockport Reservoir	0.382	0.068	0.040	0.012	175.0	71.8	1.3	12.6	49.3	17.04	116.1
Weber River	Above Oakley City POTW	0.109	0.023	0.009	0.002	173.0	72.1	1.0	12.6	28.7	12.69	35.1
Weber River	Below Oakley City POTW	0.125	0.040	0.020	0.017	175.0	71.6	1.1	20.1	33.7	11.19	35.8

Notes: POTW = publicly owned treatment works, SD = standard deviation, TN = total nitrogen, TP = total phosphorus.

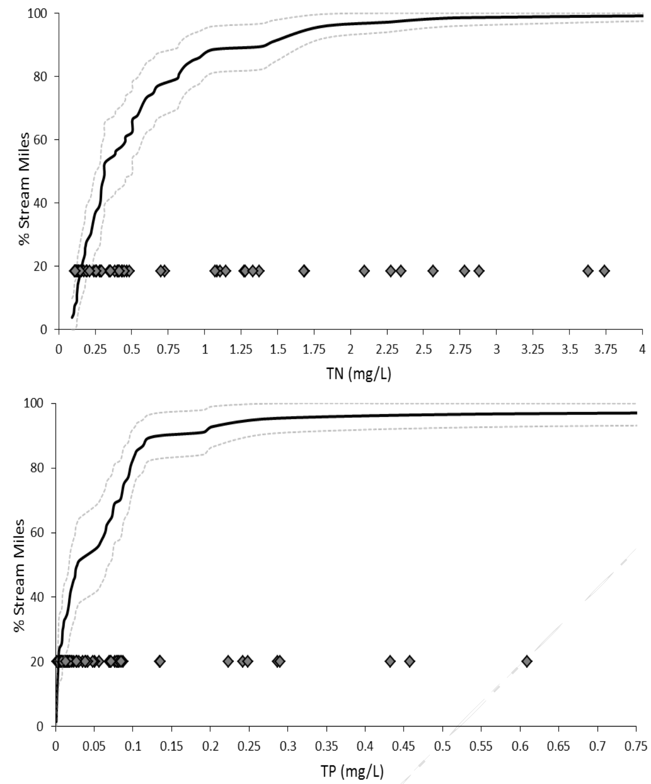


Figure 3.3. Solid black lines represent the cumulative frequency distribution from randomly selected sites throughout Utah, and dashed lines depict the 95% confidence interval of these relationships. Grey diamonds are the average nutrient concentrations obtained from the sites used in this functional response study. These plots do not include three high total nitrogen and four high total phosphorus sites because they exceeded the plot scale.

DWQ will continue to conduct more detailed covariate analyses. For instance, a more thorough evaluation of the influence of covariates on headwater N and P concentrations has already been completed in anticipation of headwater NNC development (Chapter 15), and tools that could be applied to NNC development on a site-specific basis have been identified (Chapter 16).

Chapter 4

EXPERIMENTAL ESTIMATES OF NUTRIENT LIMITATION AND SATURATION

Key Points

Nutrient-diffusing substrate experiments were conducted to determine patterns of nutrient limitation and saturation at 29 streams.

Based on the data collected from reference sites, which represent the natural state of Utah streams, primary production is limited by nitrogen or co-limited by nitrogen and phosphorus.

The presence of *both* nitrogen and phosphorus is associated with greater increases in biofilm biomass relative to controls than the presence of nitrogen or phosphorus alone.

On average, streams are more likely to be nutrient saturated once total nitrogen exceeds 0.42 mg/L or total phosphorus exceeds 0.078 mg/L, although the data show considerable variation around both thresholds.

Introduction

Within streams, nutrient availability is related to rates of autotrophic production (Bernhardt and Likens 2004, Fairchild et al. 1985), as well as whole ecosystem primary production (Bernot et al. 2010, Elser et al. 2007, Mullholland et al. 2001). This relationship has been extensively described in lakes, where the supply of nitrogen (N) and, especially, phosphorus (P) are strongly related to phytoplankton biomass (Elser et al. 1990, Hecky and Kilham 1988, Schindler 1977). Others have demonstrated the importance of N and P in predicting stream algal biomass (Biggs 2000, Chetelat et al. 1999, Dodds et al. 1997), although relationships are less generalizable. And numerous studies indicate that the growth and productivity of autotrophic biota in aquatic systems are often limited by nutrient availability, most commonly N and/or P (see Elser et al. 2007 and Francouer 2001 for reviews).

At regional scales, however, the relative effects of different water column nutrient concentrations (i.e., N:P ratios) on benthic algal growth is more complex. For instance, when comparing N:P ratios to chlorophyll-*a* (chl-*a*) measurements across multiple stream systems, Francoeur (2001) and Johnson and colleagues (2009) observed both positive and negative correlations between water column nutrients, physical stream features, and benthic algae growth (measured as chl-*a*). Important physical factors with the potential to limit aquatic primary production include light availability (Bernhardt and Likens 2004, Hill and Fanta 2008), scouring or substrate stability (Mulholland et al. 1991), and temperature (Sanderson et al. 2009). In larger soft-bedded rivers the sorption and subsequent release of nutrients from stream sediments is another important factor. In rivers with extensive macrophyte beds, nutrient uptake can be rapid and retention longer because few organisms directly feed on plant tissue (Riis et al. 2012). Other biotic interactions, such as competition with heterotrophs (Tank and Dodds 2003) and invertebrate grazing (Opsahl et al. 2003, Rosemond et al. 2000), also alter rates of primary production in stream ecosystems.

The relative importance of different controls on stream algal production is of keen interest to water resource managers. Knowing which nutrients, if any, limit algal growth helps resource managers make informed decisions about specific mitigation actions that are most likely to improve or maintain the biological integrity of aquatic ecosystems. This is particularly true for point sources with treatment facilities capable of removing either N or P. Whether or not a facility treats for N, P, or both can significantly affect operating costs (e.g., see CH2MHill 2009). Nonpoint sources are more difficult to track and control, but knowing which nutrient is limiting helps managers select the best management practices (BMPs) that are most likely to result in water quality improvements. These different nutrient sources are also potentially important in the alteration of limitation patterns, because nonpoint nutrient sources are episodic, whereas point sources are much more constant.

One way to measure nutrient limitation is with bioassays. In streams, the most common bioassay approach is the use of nutrient-diffusing substrates (NDS) (Johnson et al. 2009, Scott et al. 2009, Tank and Dodds 2003), in part because of the relatively low cost of these experiments. NDSs are composed of an agar medium infused with a single nutrient or combination of nutrients (most commonly N, P, or both) that diffuse through a solid porous substrate into the water column. The solid porous substrate serves as a platform for algal colonization and can be easily removed to measure algal biomass accrual. With these experiments, the nutrient (or combination of nutrients) treatment that leads to the greatest increase in algal biomass—typically expressed as ash free dry mass (AFDM) or chl-*a* per area of stream bed—is assumed to be the one mostly likely to limit benthic production.

In this study, Division of Water Quality (DWQ) used NDS to evaluate nutrient limitation on benthic primary production at 35 streams throughout central and northern Utah. The objectives of this study were to: (1) determine the limiting nutrient for algal growth in minimally degraded reference sites, (2) evaluate the effects of moderate and high levels of eutrophication on nutrient limitation, and (3) quantify the total N (TN) and total P (TP) concentrations where saturation with respect to NDS biomass accrual generally occurs. These three indicators provide insight into the relative importance of N and P in benthic primary production.

Methods

Study Sites

During the summer of 2010 NDS bioassays were deployed in 29 of the 35 sites in the functional response study. Of these sites, 15 were in reference condition and relatively uninfluenced by human-caused perturbations. The remaining 14 study sites were located upstream (n = 7) and downstream (n = 7) of publicly owned treatment works (POTWs) to encompass the range of ambient stream nutrient levels and the variety of nutrient sources among streams throughout Utah (see Figure 2.3, Chapter 2). Deployments took place from July to August during baseflow conditions and when potential algal growth rates were greatest due to higher temperatures and longer days.

For the purpose of NDS analyses, *a priori* site classes were developed based on differences in nutrient sources. A summary of these *a priori* site categorizations follows:

Reference Sites: Minimal effects from human activities, including nutrient inputs.
Moderately Enriched Sites: Primarily nonpoint source nutrient inputs, variable but sometimes extensive habitat modifications.
Highly Enriched Sites: Both point source and nonpoint source nutrient inputs, variable but sometimes extensive habitat modifications.

Nutrient-Diffusing Substrates

To assess nutrient limitation on algal growth, DWQ constructed NDS housing and nutrient-enriched agar using procedures developed by Utah State University's (USU) Aquatic Biogeochemistry Laboratory, which followed the methods presented by Tank and colleagues in *Methods in Stream Ecology* (2006). NDS units were constructed from 1-ounce Poly-Con® cups (Median Plastics) with a ¾-inch hole drilled into a hinged lid. The cups were filled with approximately 30 mL of agar solution with different combinations of nutrients to obtain four different nutrient treatments: control (Agar); N as NH_4Cl, 0.5 mole/L; P as KH_2PO_4, 0.5 mole/L; and N + P (in the same forms used in the single-nutrient cups). Each NDS housing unit was topped with a 2.75 cm-diameter fritted glass crucible cover (Leco, Inc) to provide an inorganic surface for periphyton to grow (Figure 4.1). To facilitate stream deployment, a housing unit was constructed for the nutrient treatments from a 12-inch plastic L-shaped bar (US Plastic, part 48445); the bar was secured to a cinder block. Each cinder block held three replicates for each of the four nutrient treatments. Cinder blocks were deployed perpendicular to streamflow in a representative glide within each stream.

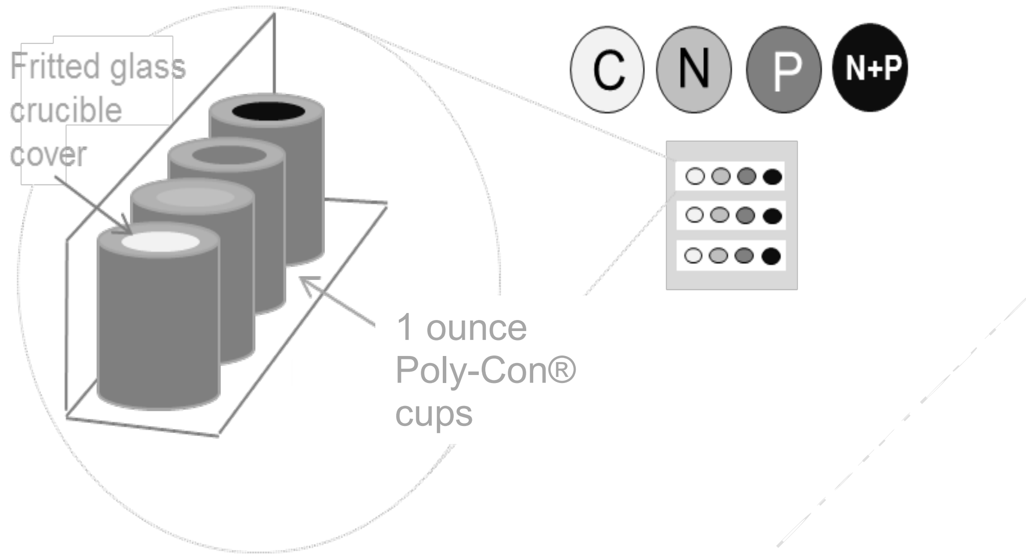


Figure 4.1. Diagram of nutrient-diffusing substrate housing and the configuration of treatments.

Deployment protocols were developed to minimize variation in other factors (i.e., shading and scour) that could potentially obscure among-stream differences in NDS patterns. At each site, a deployment location in a representative flowing water habitat (i.e., glide or run) was selected. NDS deployed upstream of POTW discharges were positioned above, but as close as possible, to the POTW discharge location. NDS deployed downstream of POTW were placed directly below the mixing zone of the effluent and receiving waters. To determine the mixing zone, the specific conductance at ≥5 points across the stream channel was measured using YSI data sondes at several locations downstream. Measurements were continued until conductivity measurements across the width of the stream were homogenous, after which the next downstream glide or run was selected as the deployment site.

Once representative glides or runs were established, a specific deployment location in an area that maximized sun exposure was selected. This helped minimize, to the greatest extent possible, among-stream differences in channel shading. NDS cinderblocks were arranged perpendicular to the flow to minimize among-treatment contamination. Following standard deployment protocols, NDS bioassays were incubated for ~21 days (Tank et al. 2006). NDS housings were inspected twice during the incubation period, and any visually observable settled organic material or entangled debris (e.g., grasses or small sediment debris) that had accumulated on the cinder block and/or NDS housing units was cleared. After the incubation period, the crucible covers were removed and immediately frozen to preserve the samples and then thawed immediately prior to laboratory chlorophyll analyses.

Water Chemistry

Surface water grab samples of 50 mL were removed from upstream and immediately downstream of the NDS three times during the NDS incubation period. Samples were kept frozen until analyzed by USU's Aquatic Biogeochemistry Laboratory for TN and TP (see Chapter 2 for details; also see Valderrama

1981). Chl-*a* analyses were conducted by the State of Utah Public Health Laboratory with standard fluorometric methods uncorrected for pheophytin (USEPA Method 445.0).

Data Analysis

Quantification of Experimental Responses

We quantified among-treatment differences in algal accrual by measuring the relative differences in chl-*a* that accumulated on each fritted glass cover during deployments. In most cases, NDS chl-*a* results (mg/m^2) are expressed as an absolute 21-day accrual. However, in several instances (n=8), deployments lasted slightly longer than 21 days due to field logistics, which necessitated use of a linear correction to standardize among stream comparisons.

A two-factor analysis of variance (ANOVA) was used to test whether chl-*a* increased on N and/or P treatments. Prior to ANOVA analyses, chl-*a* concentrations were square root (x) transformed to meet statistical normality assumptions. Following general conventions, statistical significance (alpha) was set at $p < 0.05$. At 11 sites, a replicate sample was lost; in those cases the remaining replicates were equally weighted to balance the ANOVA. The statistical outcomes of these tests were translated into ecologically meaningful nomenclature following the work of Tank and Dodds (2003):

- ***Single nutrient limitation*** occurs when N or P (but not both) elicit an increase in NDS algae accrual.
- N and P are ***independently co-limiting*** when neither N nor P treatments elicit an increase in benthic algae, but the addition of both N and P increases NDS algae accrual.
- N and P can also be ***co-limiting*** when N and P additions both elicit a positive response, even if the addition of both does not result in additional algae accrual.
- It is possible that NDS algae accrual is ***not limited by either N or P***—for instance, if production is limited by another factor (i.e., light, micronutrient). In this case, neither N nor P nor the two nutrients together elicit an increase in NDS primary production.

Saturation Thresholds

Determining whether it was possible to identify in-stream concentrations of TP and TN where nutrients were saturated (i.e., no longer limiting to algal growth) was another goal of the study. For these analyses, sites were divided into two groups: those that showed nutrient limitation and those that did not. Nonparametric deviance reduction (NDR) from Qian and colleagues (2003) package rpart with water column nutrients as the independent variable and nutrient limitation as a binary, dependent variable (e.g., any or no nutrient limitation) was used to establish N and P thresholds that best delineated these groups. To test the significance of the thresholds, a linear mixed model (package lme4) for sites on each side of the threshold was used to determine whether there were any differences in limitation. Pairwise differences between groups was subsequently evaluated with ANOVA followed by Tukey's honestly significant difference *post hoc* tests.

One potential problem with NDR approaches is that derived thresholds are sometimes misleadingly robust because they are overly dependent on the specific dataset used to generate them. To determine whether results of this study were applicable beyond the dataset used to derive thresholds, the receiver operator characteristics (ROC, package pROC) was used. ROC evaluates the performance of a binary classification system using bootstrapped data to generate several metrics of model performance. The strengths of thresholds identified by NDR were evaluated with several of these metrics, including the area under the curve (AUC), the sensitivity (true positive error estimate), and specificity (true negative error estimate) of TN and TP thresholds. AUCs quantify the accuracy of models by calculating the probability that at the threshold established a randomly chosen site has the predicted response given previously defined “acceptable” Type I and II error rates. In this study, AUCs provided the percent of the measurements for which thresholds correctly identified sites saturated by nutrients. Other measures of the performance of these threshold models included sensitivity and specificity using the following equations:

Sensitivity = True Positives/True Positives + False Negatives
Specificity = True Negatives/True Negatives + False Positives

All analyses were conducted in R v2.15.0 (R Core Team 2012).

Results

Enrichment Classes

Significant differences in TN and TP were observed among the three nutrient-enrichment classes (ANOVA, $p < 0.001$):

- Reference Sites: n = 15, TN 0.25 and TP 0.027 mg/L
- Moderately Enriched Sites: n = 7, TN 1.16 and TP 0.098 mg/L)
- Highly Enriched Sites: n = 7, TN 6.57 and TP 1.72 mg/L)

The extent to which NDS nutrient limitation varied systematically with nutrient enrichment was then evaluated using these classes of stream enrichment.

Nitrogen versus Phosphorus Limitation among Utah Streams

Among all sites, several patterns of nutrient limitation were observed: no limitation, sole N- or P-limitation, and co-limitation (Table 4.1). The most common condition observed among all sites was co-limitation of N and P, which was observed at 12 sites. The next most common pattern was nutrient saturation—no increase in chl-*a* concentrations in response to nutrient additions—which was observed at 10 sites, half of which were within the highest enrichment class. Relatively few sites (n=7) were limited by a single nutrient, and among these sites N-limitation (n=5) was more common than P-limitation (n=2).

Table 4.1. Nutrient limitation determined by nutrient-diffusing substrate deployments and ambient water column nutrient concentrations for each site in the study. Treatment effect was determined by a two-way analysis of variance. Site descriptions can be found in Chapter 3, Table 3.1.

Site Code	Nutrient	Total Nitrogen (mg/L)		Total Phosphorus (mg/L)		Site Type
	Limitation	Mean	SD	Mean	SD	
BLACKFK	N+P	0.188	0.040	0.008	0.001	Reference
DCSP_AB	None	2.216	0.895	0.135	0.063	Moderately Enriched
DCSP_BL	None	11.286	3.083	1.894	0.470	Highly Enriched
DIAFK	N	0.410	0.063	0.084	0.041	Reference
FISHCK	N+P	0.246	0.049	0.063	0.072	Reference
KIMBALL	N+P	0.279	0.057	0.028	0.004	Reference
LBRAVON	None	0.343	0.234	0.021	0.009	Reference
LBRW_AB	N	1.175	0.183	0.075	0.014	Moderately Enriched
LBRW_BL	None	1.085	0.163	0.084	0.010	Highly Enriched
LOGR1000	None	0.139	0.080	0.012	0.005	Reference
LOGRDUG	None	0.424	0.030	0.023	0.002	Reference
LOGRTB	N+P	0.132	0.025	0.012	0.004	Reference
MRTRE_AB	P	2.828	0.570	0.236	0.051	Moderately Enriched
MRTRE_BL	N+P	3.897	1.204	0.445	0.144	Highly Enriched
NFCHLK	N+P	0.161	0.040	0.006	0.003	Reference
PRICER	N+P	0.388	0.048	0.050	0.014	Reference
SCSNYD_AB	N+P	0.319	0.121	0.015	0.008	Moderately Enriched
SCSNYD_BL	None	14.717	4.613	2.212	0.715	Highly Enriched
SFKLBR	N	0.229	0.040	0.017	0.003	Reference
SPRFV_AB	None	1.324	0.215	0.019	0.009	Moderately Enriched
SPRFV_BL	None	1.677	0.232	0.078	0.035	Highly Enriched
SPRM_AB	P	1.232	0.116	0.033	0.008	Moderately Enriched
SPRM_BL	None	10.416	1.194	7.897	1.449	Highly Enriched
TIEFK	N+P	0.118	0.011	0.007	0.002	Reference
UKMURD	N	0.109	0.017	0.004	0.001	Reference
UPRNFK	N+P	0.113	0.023	0.003	0.003	Reference
WEBR	N	0.382	0.068	0.040	0.012	Reference
WROAK_AB	N+P	0.109	0.023	0.009	0.002	Moderately Enriched
WROAK_BL	N+P	0.125	0.040	0.020	0.017	Highly Enriched

Note: SD = standard deviation.

General Limitation Patterns: Nutrient Treatments versus Controls

With the exception of reference site P treatment, NDS algae accrual was, on average, predictably higher for nutrient treatments relative to controls. The magnitude of treatment responses differed among enrichment classes, with reference sites showing the greatest response followed by moderately and then highly enriched streams (Figure 4.2). N generally had greater influence on NDS algae accrual than P. This was particularly true for treatments that augmented both N and P. On average, the addition of both N and P resulted in as much as 3 times more algae accrual than the addition of P alone. While this pattern was found among all nutrient-enrichment classes, the magnitude of the response generally decreased with increasing ambient nutrient concentrations. Many of the streams with the highest background nutrient concentrations did not respond to any of the nutrient treatments.

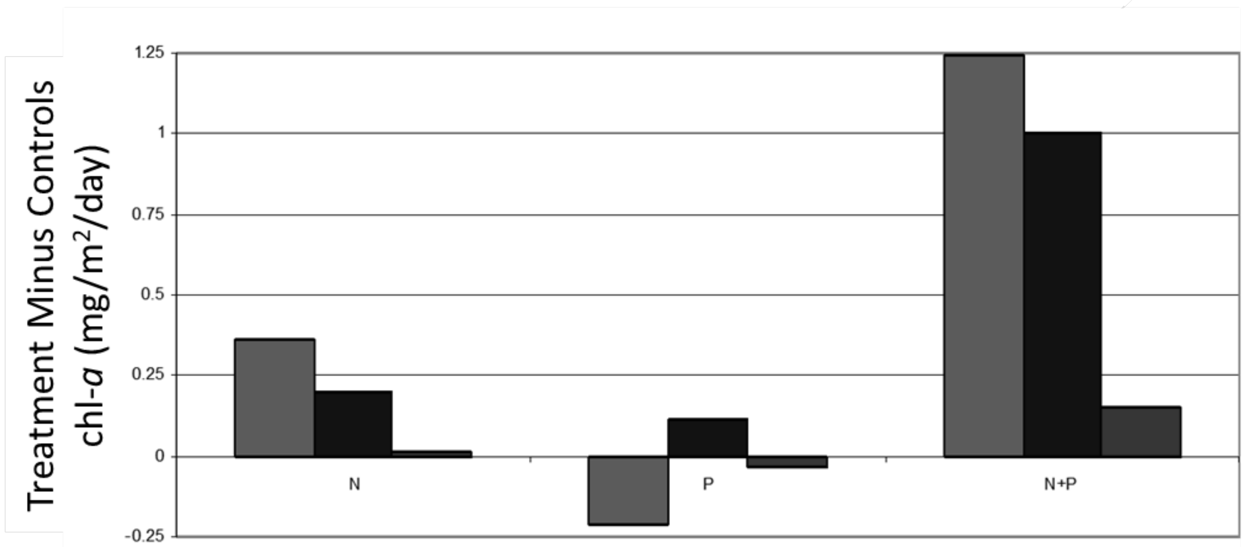


Figure 4.2. Accrual of chlorophyll-*a* for treatments relative to controls among streams within each of the three nutrient classes: reference (lightest gray), moderately enriched (black), and highly enriched (intermediate gray) streams.

Limitation Patterns among Enrichment Classes

Patterns of limitation also differed among nutrient classes. Significant increases in chl-*a* were observed in treatments relative to controls among reference sites for both N (mixed linear model, $p = 0.001$) and N and P treatments ($p < 0.001$, Table 4.2). However, enriched sites were not consistently N, P, or co-limited. When sites were binned by limitation, the only statistically distinct pattern among streams in both enrichment classes was a general tendency toward no limitation by macronutrients. In total, 71% of streams within the most highly enriched class did not respond to any nutrient treatment.

Table 4.2. Mean nutrient concentrations (± standard deviation) and the significance of the treatments within each enrichment class.

Enrichment Class	Sample Size	Mean Nutrient Concentration (mg/L)		None	N	P	N+P
		TN	TP				
Reference	15	0.25±0.12	0.027±0.025		X		X
Moderately Enriched	7	1.16±0.93	0.098±0.010	X	X*		
Highly Enriched	7	6.57±5.66	1.72±2.560	X			

Note: Statistical significance of treatments was determined with linear mixed models where X indicates significance at $p < 0.05$ and X* indicates significance at $p < 0.1$. N = nitrogen, P = phosphorus, TN = total nitrogen, and TP = total phosphorus.

Nutrient Limitation at Reference Sites

Among all reference streams, 80% were limited by one or more nutrients. The most common limiting condition found at reference sites was co-limitation, which occurred at 53% of reference streams (8 of 15). Patterns among the remaining reference sites included N-limitation, which occurred at ~26% (4 of 15 sites). Surprisingly, three reference sites (20%) were not limited by macronutrients, and no reference site was solely P-limited (see Table 4.1).

Nutrient Limitation at Enriched Sites

Limitation patterns were variable among moderately enriched streams. While most highly enriched sites did not respond to any nutrient treatment, among the streams that did, one was N-limited and two were co-limited. Sites with moderate levels of nutrient enrichment were the most variable and exhibited every type of nutrient limitation in roughly equal proportion: approximately 29% (2 of 7 sites) were limited by P, 29% by N and P, 29% did not respond to treatment, and 14% (1 of 7) were limited by N.

Saturation Thresholds

Deviance reduction analysis was used to identify ambient nutrient concentrations at sites where the N and P treatments did not increase algae growth (i.e., saturation thresholds) and at sites that exhibited an experimental response. Sites were more likely to be saturated above a TN of 0.42 mg/L, (95% confidence interval [CI], 0.33–1.4 mg/L), whereas TP saturation thresholds occurred at a concentration of 0.078 mg/L (95% CI, 0.017–1.33 mg/L). Both thresholds have been confirmed as reasonably robust with ROC models. Both models were reasonably accurate with an AUC of 81.7% (95% CI, 64.0–99.4%) for TN thresholds and 71.9% (95% CI, 51.0–92.9%) for TP, which indicates that sites would be correctly classified into these groups about 70–80% of the time. Other measures of accuracy were equally promising, with sensitivity (TN = 0.78 [95% CI, 0.44–1.0], TP = 0.56 [95% CI, 0.23–0.89]) and specificity (TN = 0.75 [95% CI, 0.55–0.90], TP = 0.80 [95% CI, 0.60–0.95]) both suggesting a balance between Type I and II errors.

Patterns of nutrient limitation among sites above and below the TN and TP saturation thresholds were evaluated; on average, sites below the thresholds show patterns similar to those observed among reference sites. Linear mixed models revealed that sites with ambient TN and TP below saturation thresholds had significant N-limitation ($p < 0.001$) and co-limitation ($p < 0.001$). In contrast, sites with ambient nutrient concentrations above both thresholds did not respond to any of the nutrient treatments, which indicates that these sites were at saturation points with respect to TN and TP (for all treatments $p > 0.37$; Figure 4.3).

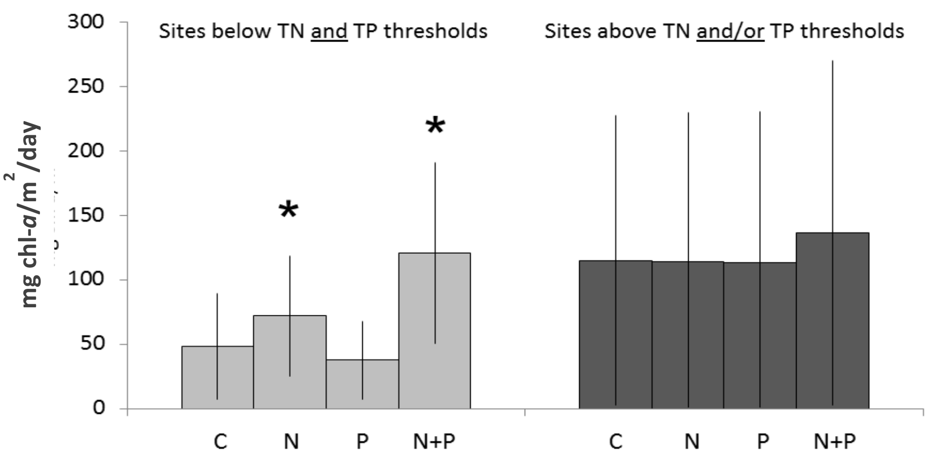


Figure 4.3. Chlorophyll-*a* concentrations by treatment for sites below the total nitrogen and total phosphorus thresholds (left panel, 0.42 and 0.078 mg/L, respectively, n = 16) or above either threshold (right panel, n = 13). Error bars are one standard deviation. Asterisks indicate a significant treatment effect using a linear mixed model.

Discussion

For more than a decade the U.S. Environmental Protection Agency (USEPA) has emphasized the importance of states developing numeric nutrient criteria (NNC) to protect beneficial uses (USEPA 2010). Development of protective nutrient criteria requires an understanding of the relationship between nutrients and the deleterious effects that nutrient enrichment can cause in a waterbody. Increased periphytic algal productivity is among several import primary responses to increased nutrients and can lead to decreased aesthetic values and issues that cause deleterious impacts to recreational uses (Suplee et al. 2009; also see Chapter 8 in this report) and/or contribute to large dissolved oxygen daily fluctuations with low nighttime minima that can be harmful to macroinvertebrates and fish (see Chapter 7 for aquatic life uses).

This study, among others (i.e., Elser 2000 meta analysis), consistently demonstrates that streams are highly variable with respect to nutrient limitation, so understanding the relative role of N and P in determining algal accrual is useful in the context of meeting nutrient reduction management objectives.

This study demonstrates the potential utility of NDS bioassays as a potentially useful tool for resource managers.

Limitations of Bioassay Experiments

When exploring stressor-response relationships, there are advantages and disadvantages to using data collected from ecological surveys and data obtained from controlled experiments. Ecological surveys are most realistic but can be difficult to interpret because many important ecological characteristics co-vary. Experiments seek to control some of these factors to study others, but ecosystem-scale experiments are expensive and difficult (if not impossible) to replicate, and the results of smaller-scale experiments (such as the NDS bioassays discussed in this chapter) cannot always be accurately extrapolated to the ecosystem of interest. Both approaches have merit, but the limits of each approach much be considered when interpreting the data.

For this investigation, the primary goal was to understand how nutrient limitation differed among streams with different ambient nutrient concentrations. As a result, the experiment controlled several other factors that influence benthic production. The experiment was conducted during the peak growing season and assumed that nutrients would be similarly limited during other periods of the year. In fact, some investigations have documented temporal shifts in nutrient limitation (Francoeur et al. 1999, Sanderson et al. 2009), which have been attributed to seasonal changes in temperature, temporal fluctuations in ambient nutrient concentrations or stoichiometry (Francoeur et al. 1999), or seasonal changes in light availability (Rosemond et al. 2000).

Another important influence on NDS nutrient limitation is light (Francoeur et al. 1999, Schiller et al. 2007). Among-stream differences in channel shading were controlled for by deploying the NDS experiments in areas of the stream with minimal channel shading, which assumes that patterns of nutrient limitation are similar under both low and high light conditions. This approach may have resulted in underestimation of saturation concentrations, but this would only be the case if many of the sites had both high nutrients and an extensive canopy cover, which rarely occurred. This approach may have also resulted in underestimation of the number of streams where light, as opposed to nutrients, is the factor most limiting benthic algae growth.

Patterns of nutrient limitation observed in this investigation may also have been altered if the species of algae growing on the fritted glass filters were different from the species of algae that grow in natural stream habitats. Algae species do differ in nutrient requirements and species-specific requirements can vary over time. However, evidence of the extent to which these compositional differences affect NDS results is mixed. In one study by McCormick and colleagues (2001), the macroalgae *Cladophora* in a stream was generally P-limited, but sometimes shifted to being N-limited due to luxury P uptake during periods of high ambient P concentrations in the stream.

While the specific design of this experiment may have influenced the results, it may not have altered the broad patterns of nutrient limitation that were observed among Utah streams (Capps et al.

2011). Even if experimental artifacts occurred, they would not necessarily negate the potential value of NDSs as a line of evidence in support of Utah's nutrient management strategy. Nevertheless, the assumptions used in this experiment will need to be assessed on a site-specific basis before concluding that the control of N or P is not needed to address a nutrient-related impairment. The relative influence of these factors could be evaluated, in part, with systematic alteration of the experimental design of NDS bioassay experiments.

Nutrient Limitation at Reference Sites

It is useful to examine patterns of nutrient limitation among reference sites because, as demonstrated in this investigation, limitation patterns can change as streams become degraded. Sites in reference condition are frequently used to establish water quality goals or restoration expectations (Stoddard et al. 2006), although reference conditions are not always indicative of the best attainable condition, particularly at streams with irreversible habitat or hydrologic alterations. Based on this NDS study, it appears that benthic algae production in Utah's reference streams is most likely to be N-limited or co-limited by N and P because these conditions were observed at 80% of all reference sites. None of the evaluated reference sites were solely P-limited.

The reference site NDS results highlight the importance of including both N and P in Utah's nutrient reduction strategy. On average, algal growth on N treatments were ~36% greater than control treatments, whereas N and P treatments were ~124% greater than the control. Additionally, the combined influence of generally increased algae responses (Figure 4.2) highlights the importance of considering both nutrients in the context of Utah's nutrient reduction strategy. Results from linear mixed models showed significant N effects and a significant N and P interaction (Table 4.2). In Utah streams, these studies suggest that it may be more appropriate to classify stream benthic algae growth as primarily N-limited and secondarily P-limited rather than as strictly N and P co-limited (Tank and Dodds 2003). At streams with these characteristics, increases in N would be predicted to increase benthic algal growth initially, but the N demands of algae would be met relatively quickly, requiring P increases for further algae accrual.

Co-limitation of N and P appears to be the predominant natural state among Utah's streams, an observation that concurs with several recent reviews of freshwater nutrient limitation that stressed the importance of co-limitation over the long-held paradigm of single-source P-limitation (Harpole et al. 2011, Lewis et al. 2011). One meta-analysis of experimental nutrient-enrichment studies found that, in general, additions of both N and P in freshwater, marine, and terrestrial environments led to a greater response than either nutrient alone (Elser et al. 2007). In addition, the magnitude of responses in freshwater ecosystems may be exponential, because the response to additions of both N and P is greater than the sum of responses to individual additions of N or P (Allgeier et al. 2011).

Although 12 of the 15 reference sites showed some form of limitation, three sites did not respond to experimental nutrient increases. These sites had low nutrient concentrations (means of 0.23 and 0.016 mg/L TN and TP, respectively) that were similar to the average overall population of reference streams (means of 0.24 and 0.025 mg/L TN and TP, respectively). Something other than N and P likely controls benthic algae growth rates in these streams. One possibility is that algal growth at these sites was limited by micronutrients such as cobalt, iron, manganese, molybdenum, or zinc (Passy 2008, Pringle et al. 1986) instead of N or P. A second, more likely, possibility is that these sites were light limited. Each site in this experiment had heavy shading. To minimize shading effects at reference sites, deployment locations that received the most direct sunlight were selected; nevertheless, site-specific shading effects at these locations may not have been completely eliminated or controlled. Therefore, the results of this experiment demonstrate the effects of nutrients in the more sensitive stream reaches where light is not the predominant limiting factor in benthic algae accrual. These results further highlight the importance of understanding the influence of site-specific covariates when interpreting ecological responses to nutrient enrichment.

Patterns of Limitation among Enriched Streams

Among those enriched streams that were not saturated with N or P, limitation patterns were highly variable. One possible explanation is that the variable responses are manifestations of human-caused enrichment. Different types of human-caused nutrient enrichment could alter stream N:P, and such changes could cause a shift from, for instance, P- to N-limitation. These variable limitation patterns could also be the consequence of among-stream differences in algae composition (Borchardt 1996), site-specific physical conditions (Rosemond et al. 2000), or the nature of human-caused nutrient inputs (i.e., pulses versus press). Temporal variation may also play a role. Streams are known to shift limitation from one nutrient to another seasonally, especially if a shift in the predominant algae taxa occurs. The extent to which algae composition changes seasonally differs among streams, which could lead to variable limitation patterns among unsaturated streams in this investigation.

Human-caused shifts in patterns of nutrient limitation have considerable management implications because the relative importance of N versus P may change along recovery trajectories. Under such circumstances, controls for both N and P are more likely to improve ecological responses. Another implication for enriched streams relates to the nutrient augmentations central to NDS experiments. A better experiment might be one that estimates limitation patterns under nutrient-removal scenarios, but this may be difficult to accomplish in situ. Another possibility would be evaluating changes in C:P and C:N in algae tissue following reciprocal transplants of cobbles from high- and low-nutrient streams. King and colleagues (2000) tested this approach under experimental nutrient additions in artificial stream channels and found that the C:P of algae from low-nutrient streams declined with increasing nutrient treatments, whereas the C:P of algae from high-nutrient streams did not change (i.e., the algae were not P-limited).

Defining enrichment classes based on the primary sources of nutrients resulted in the inclusion of a site on the Weber River that was not enriched in comparison with other highly enriched streams. This site was among the largest in the study with base flows of ~35 cfs; it also had one of the smaller POTW discharges (0.8–1.4 cfs). In this case, the small discharge did not increase downstream nutrient concentrations (Table 4.2). The comparatively low nutrient concentrations at this site may explain why it behaves differently, in terms of limitation, than the majority of sites in the highly enriched class. These results highlight the importance of not generalizing about anticipated nutrient responses based exclusively on sources (i.e., comparisons between point sources and nonpoint sources).

Saturation Thresholds

Thresholds for both TN and TP that define the ambient conditions associated with saturated conditions were identified. Used alone, these thresholds are not definitive as NNC benchmarks due to the changes in limitation patterns caused by natural and human-caused differences among streams. However, they remain an informative line of evidence. One ecologically important implication of these benchmarks relates to the relative assimilative capacity of streams. As nutrient concentrations within streams increase, the uptake velocity decreases (Earl et al. 2006, O'Brien and Dodds 2010). This means that as streams near saturation thresholds they are increasingly unable to perform the ecosystem service of nutrient retention. Further incremental increases in nutrients at saturated streams would be transported further downstream, leading to an expansion in the spatial scale of nutrient-related water quality problems.

The significance of saturation thresholds was tested using ROC analysis and by comparing limitation among all sites above and below the TN and TP thresholds. The results of ROC suggest that the thresholds predict the presence or absence of NDS nutrient limitation quite well. For TN, the true positive (sensitivity) and true negative (specificity) prediction rates were reasonably balanced (0.78 and 0.75, respectively). For TP, a different condition of fairly low sensitivity (0.56) but high specificity (0.80) prevailed. This indicates that above the TP threshold of 0.078 mg/L, the odds are nearly even that a site will not be nutrient limited, whereas below this threshold sites are more likely to be nutrient limited, as the threshold predicts.

In general, saturation threshold patterns suggest that, on average, N may be slightly more important than P in controlling algae accrual. For instance, the ROC models indicate that TN saturation thresholds are stronger predictors of nutrient limitation than TP. Similarly, at sites where ambient nutrients were below both the TN and TP thresholds, algae were primarily N-limited and secondarily P-limited. However, as emphasized below, caution should be used when drawing conclusions from these general patterns—it cannot be said with confidence that P reductions are not needed at streams with eutrophication problems. The study identified general patterns, but remediation strategies are site-specific.

Management Implications

Regional NDS studies, such as this one, provide useful measures of nutrient responses. For instance, this study identified in-stream concentrations of TN and TP that, on average, are likely to be high enough to saturate NDS algal growth. These experimental observations are consistent with others that have linked saturation to landscape-scale patterns of land use. Bernot and colleagues (2010) observed N to be at saturation and P to be near saturation at watersheds with high levels of agriculture. Others have observed similar conditions along gradients of increasing urbanization (Meyer et al. 2007). Saturation concentrations are also important in the context of nutrient remediation efforts because improvements in algae growth would not be expected until ambient nutrient concentrations fall below saturation thresholds. Moreover, once the nutrient assimilative capacity of streams is exceeded, further nutrient inputs have a greater potential to degrade downstream uses.

NDS bioassays can also help inform and prioritize nutrient reduction efforts. By understanding which nutrient limits algal growth in a system, managers can focus resources on reducing the nutrient that will have the greatest improvement on downstream water quality, instead of implementing a one-size-fits-all nutrient reduction strategy. However, resource managers should be cautious not to over-interpret NDS experimental results. While several patterns of nutrient limitation were identified, there were exceptions. Overall, these results highlight the importance of addressing, to the greatest extent possible, both N and P when implementing nutrient reduction strategies. For decades, the assumption that P limits primary production in freshwaters was a paradigm in aquatic ecology. DWQ historically established total maximum daily load (TMDL) limits exclusively for P to address nutrient-related water quality concerns. The results of this study suggest that N limits should also be included, or at the very least considered, in development of NNC and associated TMDLs for streams. Given that the addition of both N and P frequently exhibited a greater response than either nutrient alone, the simultaneous reduction of both N and P may be the most effective remediation strategy. Moreover, the observed differences in nutrient-limitation patterns among enrichment classes suggest that shifts from N to P might be expected, and vice versa, once BMPs to reduce nutrient enrichment are implemented.

Water resource managers rely heavily on water chemistry samples as the backbone of their regulatory programs. Stream nutrient-concentration thresholds above which algal growth is likely to be unlimited by nutrients (i.e., saturated) provides valuable information for stream assessment and resource prioritization. Such thresholds may prompt further studies needed to evaluate the need and efficacy for nutrient reduction efforts. N or P saturation thresholds of 0.42 mg/L TN and 0.078 mg/L TP will be used as one indicator, among others, to inform NNC development for headwaters or in an assessment context to help identify nutrient concentrations of potential concern.

Overall, NDSs provide meaningful measures of stream functional responses. Moreover, because these experiments are inexpensive, they will be used in the future as needed on a site-specific basis. Management decisions based on NDS data must also consider other site-specific observations and nutrient responses. For example, excessive benthic algal growth is not likely to be the most important cause of degraded uses in low-gradient, soft-bedded streams, nor in larger rivers where benthic algae growth is light limited. Algae-bacteria production may become increasingly decoupled in high-nutrient streams (Scott et al. 2008), which suggests that the relative importance of autotrophic or heterotrophic nutrient limitation may also differ among different types of streams. In cobble-bedded streams, where excessive benthic algae growth is more likely to be an important nutrient response, other factors may also be important and immediate stressors to stream biota. For instance, in streams with low nighttime dissolved oxygen concentrations, heterotrophic responses to nutrient inputs may be a more important factor in the degradation of aquatic life uses. Streams that are slow-moving and depositional in nature are more complicated because they are inherently characterized by large changes in habitat, particularly increases in the number of depositional zones. This results in accrual of unstable smaller sediment particles and increased deposition of both allochthonous and autochthonous organic debris. By nature, such stream habitats support different biota than those with greater velocity and fewer depositional zones. Despite these limitations, NDS experiments certainly have an important place in the tool box of techniques that can be used to quantify ecological responses to nutrient enrichment.

Chapter 5

STREAM METABOLISM

Key Points

Daily rates of gross primary production (GPP) and ecosystem respiration (ER) obtained from whole stream metabolism models were explored as potential responses to nutrient enrichment using stressor-response relationship models.

Both GPP and ER show a statistically significant positive relationship with ambient total nitrogen and total phosphorus concentrations in study streams.

Daily rates of GPP and ER (g O_2/m^2/day) were used to place streams into good (GPP < 6, ER < 5), fair (GPP 6–10, ER 5–9), or poor (GPP > 10, ER > 9) condition classes.

For total phosphorus, a concentration of 0.02 mg/L best differentiates streams in good condition from streams in fair condition classes. A concentration of 0.09 mg/L best differentiates fair and poor condition streams.

For total nitrogen, a concentration of 0.24 mg/L best differentiates streams in good condition from streams in fair condition classes. A concentration of 1.28 mg/L best differentiates fair and poor condition streams.

Streams with high slope (> 1%) and high channel shading (> 11%) are less likely to exhibit GPP and ER responses to nutrient enrichment.

Condition classes align with Utah's dissolved oxygen criterion. An increasing number of observations fell below the criterion in fair and poor condition streams. Almost no excursions below the dissolved oxygen criterion were observed until the lower ER threshold was exceeded.

Introduction

Whole stream metabolism quantifies two ecosystem functions: ecosystem-scale rates of photosynthesis (gross primary production, GPP) and respiration (ecosystem respiration, ER). Both functions are fundamental to understanding stream ecosystems because they quantify the amount of energy that is produced (GPP) and used (ER) in the stream. Whole stream metabolism, and the calculation thereof, is based on the premise that changes in dissolved oxygen (DO) concentrations—from daytime highs to nighttime lows—are the result of photosynthesis (biologic production of O_2), respiration (biologic consumption of O_2), and reaeration (bidirectional atmospheric exchange, Figure 5.1). During daylight hours oxygen enters the water column through the physical process of reaeration and the biological process of photosynthesis (GPP). At night, DO concentrations decline because photosynthesis cannot occur in the dark. As a result, nighttime DO concentrations are strictly a function of ongoing DO gains from reaeration and losses from ER. The rates of DO increases in the daytime and losses of DO at night allow daily GPP and ER rates to be estimated from models that account for the physical determinants of DO (e.g., saturation, atmospheric pressure, temperature).

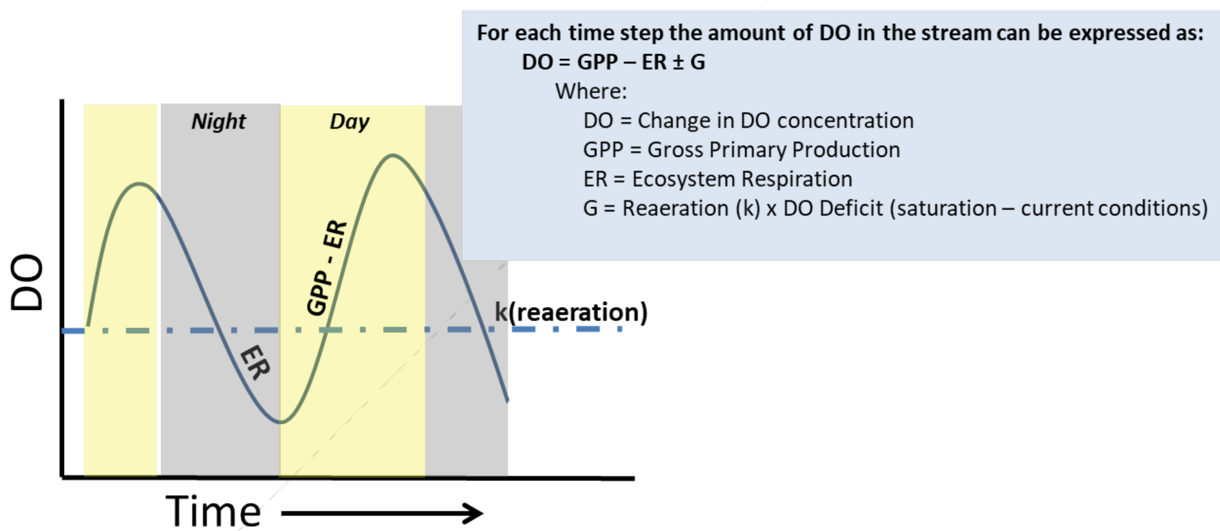


Figure 5.1. Conceptual model that depicts how ecosystem respiration and gross primary production relate to daily fluctuations in dissolved oxygen concentrations.

Researchers have used stream metabolism to investigate rates of GPP and ER since the pioneering work of Odum (1956). Since that time stream metabolism has primarily been investigated using two techniques: mesocosm (bottles or chambers) experiments and in situ—or "open-channel"—methods. Recently, open-channel techniques have gained widespread acceptance for several reasons. First, underlying data required for metabolism calculations are now reasonably accessible due to the availability of high-quality, relatively low-cost DO sensors and data loggers. Second, open-channel methods more accurately reflect reach-scale conditions because they do not introduce container effects. Finally, these techniques avoid scaling problems that can arise when extrapolating mesocosm results to stream reaches, which is the scale of interest to resource managers. Open-channel metabolism is sometimes call "whole stream" metabolism because it integrates all the metabolic processes and surface water-benthos interactions that occur over an entire stream reach (Izagirre et al. 2008, Young et al. 2008).

Aquatic ecologists have investigated both natural and anthropogenic landscape-scale influences on whole stream metabolism, such as geography (Bernot et al. 2010, Hill et al. 2000), land use practices (Houser et al. 2005, Young and Huryn 1999), and riparian disturbance (McTammany et al. 2007). Others have investigated how stream metabolic rates influence ecological processes, such as nutrient processing (Hall and Tank 2003) and ecosystem structure (Sabater et al. 2002). Together, these studies, among others, show whole stream metabolism has the potential to be an excellent indicator of stream condition. Both GPP and ER integrate several reach-scale factors that influence stream health: geomorphology, hydrology, riparian vegetation, in-stream vegetation, climate, biology, and chemistry (Grace and Imberger 2006, Mulholland et al. 2005, Young et al. 2008). On the other hand, many of these same factors vary naturally, so condition measures will need to decouple natural variation from human-caused variation for metabolism to be useful for assessing support of aquatic life uses.

This chapter explores the potential for whole stream metabolism to be used as a functional indicator of nutrient enrichment for Utah's streams. Daily rates of GPP and ER are compared to ambient nutrient concentrations across streams that have undergone a broad range of nutrient enrichment. The extent to which these relationships were influenced by several potentially important covariates (e.g., stream slope, shading and turbidity) is evaluated. To relate these relationships to protection of aquatic life uses, the relationship between daily rates of GPP and ER and Utah's existing regulatory DO water quality criteria is explored.

Methods

Data Collection

DO measurements and other data necessary for whole stream metabolism model construction were obtained at 49 stream segments across the 35 study sites (see Figure 3.1, Chapter 3), although data collected at 5 sements were subsequently excluded from further metabolism analysis due to their extremely high turbidity (details below). This left 44 metabolism observations, about half of which (15 locations) were in reference condition. A water quality probe (YSI 6600V2 or 600 OMS V2) was deployed at each site to measure DO (see Appendix A for the standard operating procedures) and temperature at 5-minute intervals for a minimum of 48 hours. Solar radiation data were obtained from the closest available weather station (mesowest.utah.edu). Surface water nutrients were collected at the time of sonde deployment and retrieval and were analyzed for total nitrogen (TN) and total phosphorus (TP) at the Aquatic Biogeochemistry Laboratory at Utah State University (Valderamma 1981).

Construction of Metabolism Models

Stream metabolism was calculated using an open water method with reaeration (*K*) as a free parameter based on the following equation derived from Van de Bogert and colleagues (2007):

$$\boldsymbol{O_t = O_{t-1} + \left(\frac{GPP \cdot \Delta t}{z} \times \frac{Light_t}{\sum Light} + \frac{ER \cdot \Delta t}{z} + K(O_{sat} - O_{t-1}) \cdot \Delta t\right)}$$

Where,

ER = Ecosystem respiration (loss of grams of O_2/m^2/day)
GPP = Gross primary production (grams of O_2/m^2/day)
K = Reaeration coefficient (day-1)
Light = Solar radiation or photosynthetic active radiation (PAR)
O = Dissolved oxygen (mg/L)
Osat = Oxygen saturation (mg/L)
t = Time (fraction of day)
z = Mean stream depth (m)

This metabolism model predicts GPP and ER at each time step to fit the oxygen data using nonlinear minimization (R function nlm) of the maximum likelihood accuracy estimates. In this equation, K can be modeled as a free parameter from the oxygen data simultaneously with GPP and ER. In rare cases where K could not be modeled accurately, K was constrained with values calculated from night time regression (Grace and Imberger 2006) to improve model performance.

Comparison of Gross Primary Production and Ecosystem Respiration to Stream Nutrient Concentrations

Generalized linear models (GLMs) were used to evaluate the relationship between average ambient nutrient concentrations (TN and TP) and the metabolic response rate indicators GPP and ER observed at each site. All data were natural log transformed prior to conducting this analysis to better reflect statistical assumptions of normality.

To further explore the relationships among nutrients and metabolism metrics, sites were classified into three groups with low, medium, and high ambient nutrient concentrations for both TN and TP. Three groups were used to be consistent with Division of Water Quality (DWQ) and U.S. Environmental Protection Agency (USEPA) assessment methods, which generally assess waterbodies in three condition classes (DWQ 2014). Two thresholds were identified; these were expected to provide ecologically meaningful information, with the first threshold corresponding to a departure from the range of natural (reference) conditions, and a second, higher threshold representing an appreciable alteration to GPP or ER processes.

To create the nutrient groups, among-group nutrient-concentration boundaries were defined using pairwise relationships between each nutrient (TP, TN) and each response (GPP, ER). Thresholds were established from each of the pairwise relationships using nonparametric deviance reduction (NDR; Qian et al. 2003). This statistical approach is computationally similar to tree-based models (Breiman et al. 1984). This analysis (R package rpart) incrementally evaluates all possible nutrient thresholds to find the concentration where the responses (GPP, ER) among the streams above and below the split are as homogeneous as possible, while also maximizing differences between the groups. The first split identified using this method established the boundary for low and medium nutrient groups. The procedure was then repeated to identify a second threshold for the moderate to high nutrient groups. The strength of each nutrient threshold was also evaluated using bootstrapped resampling methods. Analysis of variance (ANOVA) was used to determine whether the resulting nutrient groups were statistically distinct. This was followed by post-hoc Tukey's honestly significant difference (HSD) to determine whether each GPP or ER group was statistically distinct ($p < 0.05$) or any observed differences were primarily the result of one group being distinct from the others.

Differences in GPP and ER rates among the nutrients groups were also explored. Three groups with low, medium, and high metabolic response rates were created by calculating the mean TN and TP concentrations for each nutrient group, and then averaging the values obtained for TN and TP to define group splits (thresholds). This process assumes that GPP and ER concentrations associated with the lowest group reflect the range of background (i.e., reference) GPP and ER rates.

Comparisons to Dissolved Oxygen Numeric Criteria

Estimates of GPP and ER thresholds developed from this study are statistical and do not necessarily translate to the health of stream ecosystems. To provide a more direct linkage between the identified thresholds and the support of aquatic life uses, the association between GPP and ER groups and excursions of Utah's DO water quality criteria were evaluated (Utah Administrative Code [UAC] R317-2-14). These comparisons were made using two different duration intervals defined by Utah's DO criteria—the daily minimum when early life stages are not present and the 30-day average; these duration intervals represent acute and chronic threats to stream biota. The extent of oxygen criteria violations was compared among the three GPP and ER groups using ANOVA ($p < 0.05$) and then a *post hoc* Tukey's HSD test.

Because problems with low DO conditions are almost always associated with excess respiration, the relationship between ER and the number of excursions of the 30-day criterion observed during sonde deployments was explored further using GLMs. The proportion of excursions below Utah's 30-day DO standard was modeled as a nonparametric function of ER, assuming a quasi-binomial distribution for the response to account for the fact that values of the response variable were bounded by 0 and 1.

Table 5.1. Relative influence of nutrients and covariates on rates of gross primary production (GPP) and ecosystem respiration (ER). Variables with a larger increase in mean squared error (MSE) are more important determinants of metabolic rates relative to others.

Environmental Gradient	Units	GPP % Increase MSE	ER % Increase MSE	Source
Total Nitrogen*	mg/L	72.6	69.0	USU ABL
Total Phosphorus*	mg/L	73.9	52.1	USU ABL
Turbidity	NTU	41.7	37.0	UPHL
Total Suspended Solids	mg/L	50.0	35.2	UPHL
Channel Shading	%	104.3	69.3	DWQ
Slope	%	103.1	63.4	USGS
Basin Area+	mi2	58.1	30.4	USGS
Herbaceous Upland+	%	69.1	25.7	USGS
Forested Watershed+	%	67.5	32.1	USGS
Basin Slope+	%	75.3	42.3	USGS
Mean Water Depth	cm	52.6	52.6	DWQ
Mean Thalweg Depth	cm	32.7	28.0	DWQ
Bank-full Height	cm	69.4	47.3	DWQ
Channel Incised Height	cm	58.6	13.7	DWQ
Channel Width: Depth	ratio	19.2	30.5	DWQ
Fine Substrate (< 2 mm)	%	41.2	19.0	DWQ
Small Sediment (< 16 mm)	%	64.5	41.9	DWQ
Median Substrate Particle Size	cm	71.6	46.9	DWQ
Riffle and Rapid Channel Units	%	38.2	32.5	DWQ
Riparian Corridor Bare Ground	%	28.0	25.7	DWQ

Notes: Nutrients (*) are included for comparison purposes. Environmental gradients (covariates) include field measurements taken at each stream and watershed-scale attributes (+). Results were generated from random forest models that predict GPP and ER from all variables in the table. Mean squared error (MSE) is a measure of the increase in model error that results when each variable is randomized among sites, while keeping all other variable constant. Data sources include: Utah State University Aquatic Biogeochemistry Laboratory (USU ABL), Utah Unified Public Health Laboratories (UPHL), U.S. Geological Survey Stream Stats program (USGS) or the DWQ Comprehensive Assessment of Stream Ecosystems program.

Evaluation of Potential Covariates

Rates of GPP and ER vary naturally, so the relative influences of nutrients and other natural environmental gradients were also evaluated. Multivariate random forests were used (Breiman 2001; R package randomForest) to model ER and GPP from 20 potential explanatory variables that capture, directly or indirectly, characteristics that are known to—or have been suggested to—control whole stream metabolism. Potential explanatory variables were obtained from GIS data (U.S. Geological Survey StreamStats), on-site physical habitat surveys (USEPA 2009), and water quality samples (Table 5.1). In some cases, general landscape-level surrogates of stream size characteristics (e.g., basin area, water depth, width:depth) were used instead of direct measures of physical characteristics (i.e., temperature and substrate composition). This approach was used for three reasons. First, these stream characteristics generally vary systematically from headwaters to larger streams, so these landscape-scale characteristics simultaneously capture several important stream characteristics. In addition, landscape-level characteristics integrate steam characteristics, both spatially and temporally, at scales that are more closely aligned with the scale of these regional analyses. For example, at most sites only about three days of temperature data were available. Whatever summary statistic of temperature was selected for this limited data collection window could not capture ecologically relevant among-stream differences in the thermal regimes. Finally, pragmatic management applications of this work were considered, such as whether landscape-scale relationships could be applied on a statewide scale.

Random forest regression was used on all variables. The best-performing variables were then selected based on percent increase of mean squared error (MSE) that occurred under model runs where, one-by-one, the observations for each variable were randomly assigned to another stream while other variables remained constant. The assumption underlying MSE variable selection is that the most important variables will result in the largest decrease in model accuracy when the observations for that variable are randomly reassigned among sites. The magnitude of change in model accuracy provides an estimate of variable strength that is internally consistent with the model.

After selecting the most important factors that influenced GPP and ER, the analyses were re-run to create final models that reflected a reasonable balance between model accuracy and parsimony. For instance, if the best subset of variables performed as well as all variables in the random forest model (based on the pseudo-r^2 fitness statistic) then the best subset model was considered successful. The goal of this exercise was to identify important covariates that could potentially obscure or exaggerate the role that nutrients play in determining GPP or ER rates. All analyses were conducted in R v2.15.0 (R Core Team 2012).

Results

Early exploratory analyses revealed that metabolism rates were suppressed at highly turbid sites as were relationships between nutrients and rates of GPP and ER. Distributions of turbidity data revealed five highly turbid outliers with turbidities greater than 75 nephelometric turbidity units (NTU). These five highly turbid streams were subsequently excluded from the remainder of the analyses. The remaining 44 streams still encompass a broad nutrient gradient (TN 0.10–14.37 mg/L and TP 0.002–7.65 mg/L), and exclusion of the highly turbid streams should not overly bias the remaining analyses.

Relationships between Metabolism and Nutrients

GLMs revealed significant relationships between nutrients (TN and TP) and functional responses (GPP and ER) across all nonturbid sites (Figure 5.2). GPP was positively related to both TN ($r^2 = 0.30$, $p < 0.001$) and TP ($r^2 = 0.38$, $p < 0.001$). Correlations between ER and nutrients were slightly stronger for both TN ($r^2 = 0.48$, $p < 0.001$, Figure 5.1C) and TP ($r^2 = 0.50$, $p < 0.001$, Figure 5.1).

NDR thresholds for GPP and ER yielded similar, but not identical, TN and TP thresholds. The lower thresholds for TN were identical for GPP and ER (0.24 mg/L), as were the upper thresholds for TP (0.09 mg/L). In contrast, the upper TN threshold was higher for ER (1.68 mg/L) than GPP (1.28 mg/L). The

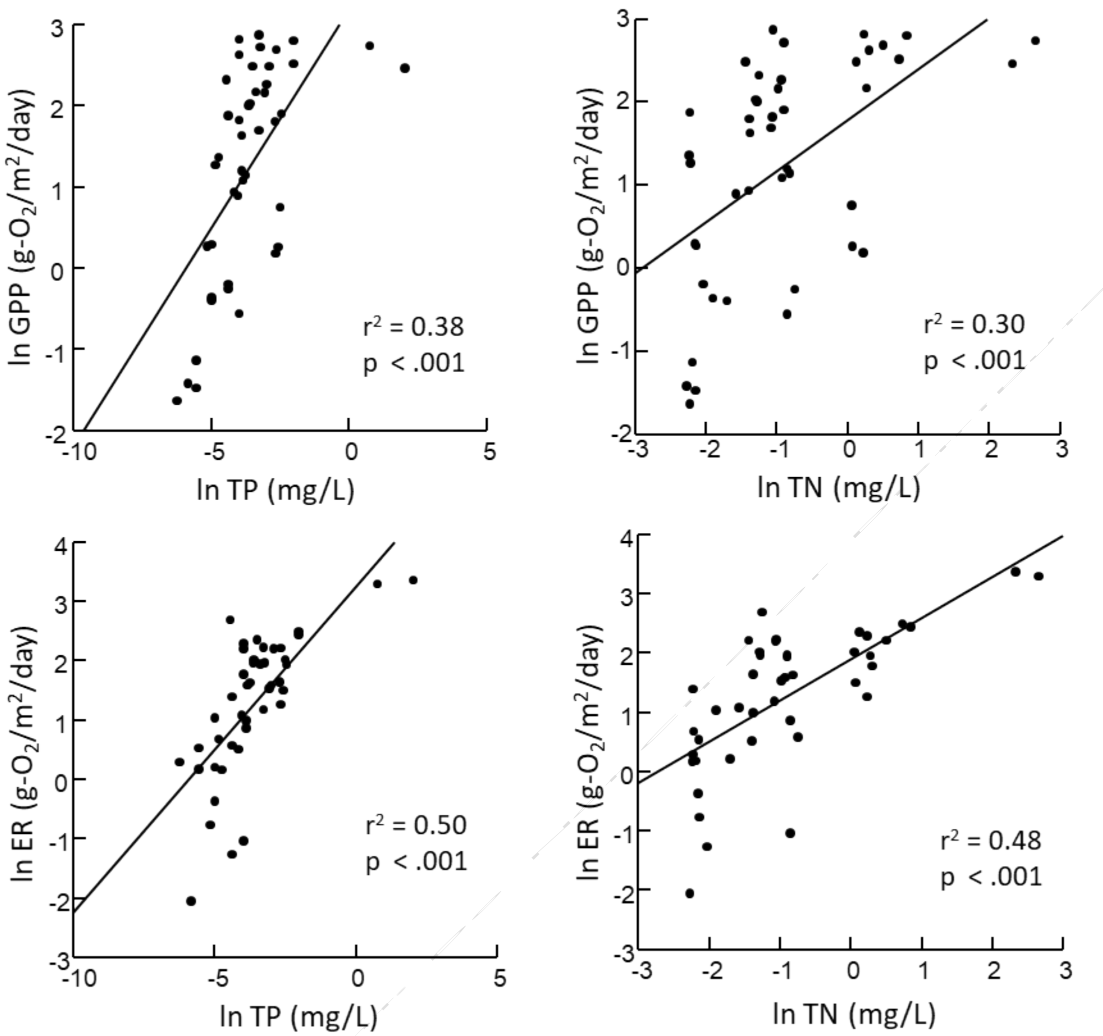


Figure 5.2. Linear relationships between ambient nutrient concentrations (total nitrogen [TN] and total phosphorus [TP]) and gross primary production (GPP, top row) and ecosystem respiration (ER, bottom row).

opposite was true for the lower TP thresholds: GPP (0.02 mg/L) was slighting higher than the corresponding threshold for ER (0.01 mg/L). Although, in this case, both were within the range of nutrient concentrations associated with headwater streams in reference condition. To simplify the interpretation and application of these data, the nutrient thresholds associated with GPP were selected to reflect general responses of stream metabolic rates to nutrient enrichment. This simplifying assumption seemed logical because high ER rates generally result from the decomposition of higher accumulations of algal biomass caused by high GPP rates.

From the deviance reduction thresholds, three distinct groups of streams were identified that were similar with respect to the mean and variance of TN and TP observations (hereafter low, medium, and high nutrient groups; Table 5.2). These statistical analyses identified the same thresholds separating low, medium, and high rates of GPP and ER: TN values of 0.24 mg/L and 1.28 mg/L, and TP values of 0.02 mg/L and 0.09 mg/L.

Table 5.2. Ambient nutrient-concentration thresholds used to place streams into relative nutrient-concentration classes.

	Low	Medium	High
Total Nitrogen (mg/L)	< 0.24	0.24–1.28	> 1.28
Total Phosphorus (mg/L)	< 0.02	0.02–0.09	> 0.09

These TN and TP nutrient groups generally corresponded predictably with measures of stream metabolism. Broadly, GPP and ER rates were significantly different among the three nutrient groups for both TN (ANOVA $p < 0.001$) and TP (ANOVA $p < 0.001$) (Figure 5.3). However, *post hoc* investigations revealed that these differences were largely a result of the low-nutrient streams being statistically distinct from other nutrient groups. All three nutrient groups were only statistically distinct for comparisons of ER with TN (Figure 5.3).

GPP differed predictably among nutrient groups with higher rates generally corresponding to sites with higher nutrients. For TN, among-group mean GPP rates (g O_2/m^2/day) were 2.43 ± 3.27 (standard deviation) for the low N group, 6.57 ± 4.9 for the medium group, and 13.19 ± 2.59 for the high group. GPP rates were similar for the TP nutrient groups: low = 3.62 ± 4.74, medium = 7.48 ± 4.75, and high=13.86 ± 2.29 (Figure 5.3).

ER also differed predictably among TN groups. From low to high, mean ER rates (g O_2/m^2/day) among the TN groups were 2.05 ± 2.28, 5.78 ± 3.29, and 14.35 ± 9.35. Mean ER rates for the low, medium, and high TP groups were 3.13 ± 3.81, 6.05 ± 2.31, and 19.66 ± 9.25, respectively (Figure 5.3).

Stream Metabolism Groups

NDR thresholds were established independently for GPP and ER to generate good, fair, and poor condition classes for each metric (Table 5.3). For GPP, a threshold of 6 g O_2/m^2/day distinguishes between good and fair condition classes, and another threshold of 10 g O_2/m^2/day distinguishes between fair and poor condition classes. The two thresholds for ER were a little lower at 5 g O_2/m^2/day and 9 g O_2/m^2/day. All groups had statistically significant (ANOVA $p < 0.05$) differences in metabolic rates.

Table 5.3. Relative condition classes for streams as defined by stream metabolism rates.

	Good	Fair	Poor
Gross Primary Production (g O_2/m^2/day)	< 6.0	6.0–10.0	> 10.0
Ecosystem Respiration (g O_2/m^2/day)	< 5.0	5.0–9.0	> 9.0

Relationships among Metabolism Metrics and Dissolved Oxygen Criteria

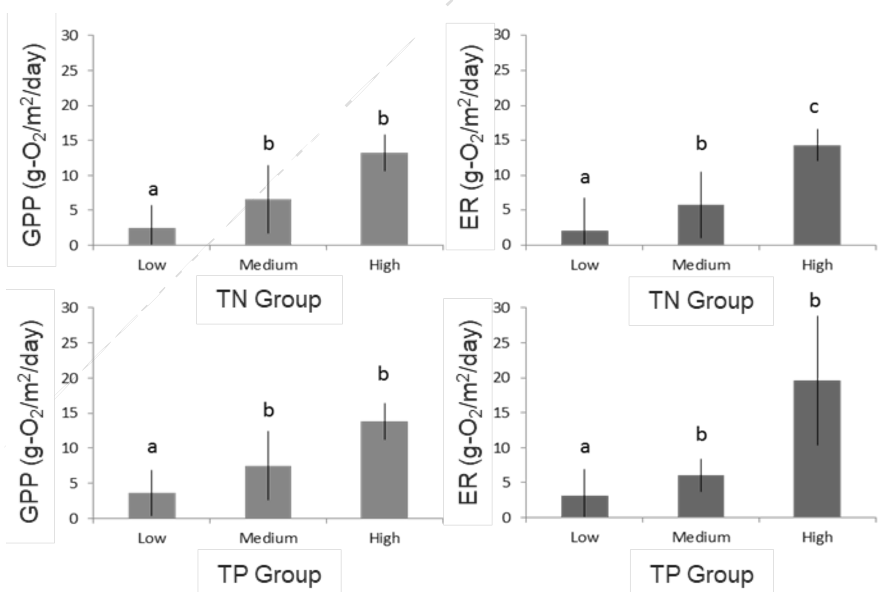


Figure 5.3. Bar charts comparing daily rates of gross primary production (GPP, light gray) and ecosystem respiration (ER, dark gray) among low, medium, and high concentration sites for total nitrogen (TN) and total phosphorus (TP). Specific group thresholds are shown in Table 5.2. Bars with different letters are significantly different from one another (Tukey's honestly significant difference, $p < 0.05$).

The extent to which the metabolic condition classes were associated with excursions below several DO criteria were evaluated with varying averaging periods because these are independently derived indicators of potential threats to stream biota. The minimum daily DO concentration was significantly different among the three nutrient groups for both GPP and ER (ANOVA, GPP $p < 0.001$ and ER $p < 0.001$). The absolute minimum DO observed during sonde deployments was generally higher at sites in the good metabolic groups.

The metabolism functional responses also corresponded predictably to Utah's DO criteria for both absolute minimum (acute) and 30-day average (chronic) averaging periods. Significant differences were identified among GPP and ER groups with regard to the relative frequency that samples at each site fell below minimum DO water quality criteria (ANOVA, GPP $p < 0.001$ and ER $p = 0.018$). For GPP, samples that fell below acute and chronic DO criteria were identified less often for sites in the good metabolic groups (Tukey's HSD, $p < 0.05$) than for sites in the fair or poor groups. The same tests revealed similar patterns among the ER groups. On average, DO observations at sites in the poor GPP and ER class fell below the acute (minimum) DO criterion ~6% of the time and the chronic (30-day) criterion ~45% of the time. These general trends obscure important site-specific differences. The fair and poor GPP and ER groups had many sites that fell below the DO criteria and many sites that did not, which led to large within-group variation in the number of excursions below the minimum DO criteria (Figure 5.4).

The GLM model used to examine the relationship between ER and excursions below the 30-day DO criterion provided useful insights. Overall, 66% of the observed variance in the proportion of DO excursions below this criterion was explained by ER (Figure 5.5). A steep increase in the proportion of excursions of this criterion was observed between approximately ER = 5 and 10 g O_2/m^2/day.

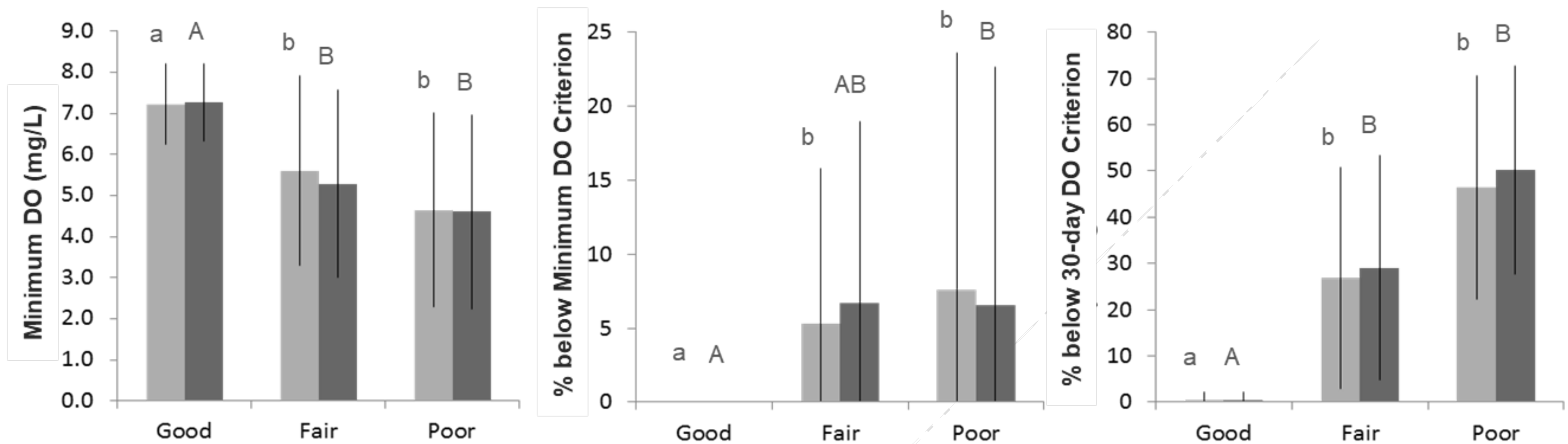


Figure 5.4. Comparisons of three measures of oxygen dynamics for three groups of streams with varying daily rates of gross primary production (GPP, light gray bars) and ecosystem respiration (ER, dark gray bars). These are the good, fair, and poor categories from Table 5.2. Bars with different lowercase letters indicate GPP groups that are significantly different from one another, and uppercase letters indicate significant differences of ER groups determined by an analysis of variance and *post hoc* Tukey's honestly significant difference test. Error bars are standard deviation.

This page intentionally left blank

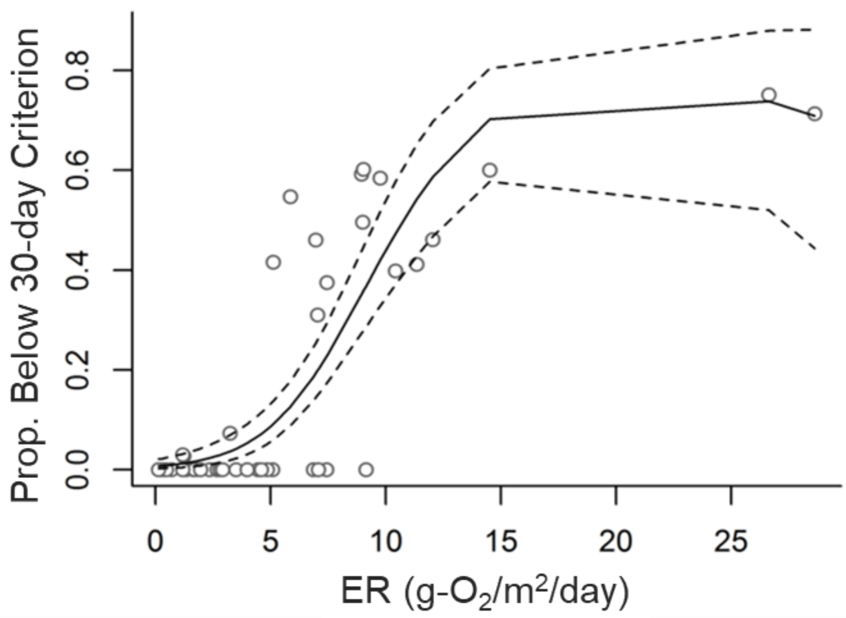


Figure 5.5. Relationship between ecosystem respiration (ER) and the proportion of site dissolved oxygen (DO) observations that fell below Utah's 30-day average DO criterion.

Physical Covariates

Random forest regression models were run separately for GPP and ER against nutrients and 20 potential covariates (Table 5.1) obtained from water quality samples, GIS analyses, and site-specific habitat metrics. The models found significant relationships between all variables and GPP (mean squared residuals = 12.7, pseudo r^2 = 0.54) and all variables and ER (mean squared residuals = 18.3, pseudo r^2 = 0.45). Four of the top five predictor variables were the same for GPP and ER, as measured by increases of MSE. The top predictor variables for GPP were stream slope (MSE=103.9), stream shading (103.3), basin slope (74.9), TN (73.4), and TP (72.6). Similar variables were most important for predicting ER, including shading (MSE = 70.2), TN (68.9), stream slope (63.2), mean stream depth (53.8), and TP (51.4). Random forest regression was run again using only the top four variables for GPP and ER (stream slope, shading, TN, and TP) to compare overall model performance. The model performed just as well with only the top four variables for GPP (mean squared residuals = 13.1, pseudo r^2 = 0.53) and ER (mean squared residuals = 16.2, pseudo r^2 = 0.51).

The relationships between ER and GPP and channel slope and shading were explored further to identify thresholds that could potentially be used to modify ER and GPP expectations, potentially increasing the accuracy of metabolism assessments. NDR revealed significant thresholds at ~1% slope for both ER and GPP. GPP and ER thresholds were also found for percent channel shading, where streams with channel shading less than ~11% had greater mean daily rates of GPP (9.3 ± 5.6 to 3.99 ± 4.1) and ER (8.10 ± 5.5 to 4.31 ± 4.1).

Discussion

Nutrient Thresholds

Daily rates of GPP and ER identified two thresholds of TN and TP that can be used to demarcate concentrations where nutrient enrichment generally alters stream metabolic functions (Table 5.2): a lower threshold that DWQ interprets as the point of departure from naturally occurring metabolic conditions and an upper threshold that reflects modifications of GPP and ER that are large enough to reflect degradation of aquatic life uses. TN values of 0.24 mg/L and 1.28 mg/L and TP values of 0.02 mg/L and 0.09 mg/L separate low, medium, and high rates of both GPP and ER. DWQ and collaborators can use these thresholds in combination with those obtained from other functional and structural indicators to identify nutrient concentrations where nutrients are high enough to affect stream conditions.

Comparison to Numeric Dissolved Oxygen Criteria

One way in which excess nutrients cause deleterious effects to stream biota is through alteration of diel oxygen dynamics via increased autotrophic or heterotrophic productivity. Stream metabolism provides ideal metrics to evaluate some of those effects because it directly quantifies the biological processes responsible for alterations to DO dynamics. Statistically binning daily rates of metabolism into three categories demonstrated significant differences among the absolute minimum DO observed at each site and percent of times that DO observations were lower than minimum DO criteria (Figure 5.3), which provides evidence that these conditions can be directly tied to independent measures of designated use support.

DWQ established DO criteria to protect aquatic life for three beneficial uses: cold water fisheries (3A), warm water fisheries (3B), and nongame fish fisheries (3C) (UAC R317-2-6). Each of these beneficial uses has a different minimum DO criterion based on differing sensitivity of fish and of other organisms in their food web. Minimum DO water quality criteria quantify, given assumptions inherent in DO standards development methods, conditions where short-term exposure is a potential threat to stream biota. Among all sites in this study, only 11% had a violation of the applicable minimum DO criterion. None of these sites had > 10% of the observations below this criterion, which is the excursion frequency that DWQ currently uses for assessments purposes. It is also true that the DO excursions observed in this study should not be confused with a violation in water quality standards, because DWQ continues to interpret DO water quality standards independently. What can be concluded from these observations is that the derived metabolism metrics, particularly ER, are related to this independent, albeit related, indicator of the condition of aquatic life uses.

Exposure to chronic DO conditions was explored by examining the percent of daily minimum DO observations for each site that fell below the appropriate 30-day average DO criterion assigned to each site and found that sites in the poorest condition, as measured by metabolism metrics, exceeded this criterion 45% of the time. For ER, a threshold response was also observed where the proportion of observations that exceeded the chronic DO criterion was much more likely to occur at ER rates greater that the lower threshold derived in this investigation. These short-term observations (48–72 hour) are not representative of actual 30-day averages, and these estimates may over- or underestimate chronic DO exposure. Nevertheless, this analysis demonstrates clear connections between GPP and ER response metrics and designated use support. DWQ currently uses the 30-day average for assessment purposes because this value is assumed to be more reflective of long-term conditions. DWQ is currently evaluating alternative methods for using instantaneous DO measurements for assessment purposes. At a minimum, these data suggest that sites with atypically high rates of summertime GPP and ER warrant follow-up investigation to determine whether low DO is also a concern.

Gross Primary Production and Ecosystem Respiration Thresholds

The thresholds of stream condition for GPP (6.0 and 10.0 g O_2/m^2/day) and ER (5.0 and 9.0 g O_2/m^2/day) developed in this study are similar to the suggested rates of GPP and ER proposed by Young and colleagues (2008) as indicators of river health in New Zealand rivers (7.0 and 9.5 g O_2/m^2/day for GPP and ER, respectively). The independently derived New Zealand metrics were obtained from a meta-analysis of metabolism calculations obtained from numerous reference sites over a 16-year period of record. The stressor-response approach used here, along with the reference condition approach developed by Young and colleagues, are part of the growing literature that provides general guidelines about the GPP and ER rates that are reflective of healthy conditions. However, the results of the random forest models also highlight the importance of taking natural changes in stream conditions into account before universally applying thresholds to infer stream condition.

Physical Covariates

This study demonstrated that nutrients were unrelated to metabolic rates at sites where turbidity was greater than 75 NTUs, which likely stems from a lack of light reaching autotrophic benthos. This indicates that stream metabolism is not an appropriate functional indicator for these sites. Nevertheless, at these highly turbid sites the mean TN (2.41 mg/L) and TP (0.36 mg/L) concentrations were an order of magnitude greater than the highest numeric nutrient criteria proposed elsewhere, so other indicators are likely to detect nutrient-related impairments at these sites. Moreover, these highly turbid streams are atypical because streams with > 75 NTU constitute less than three percent of the total stream miles in Utah (DWQ, unpublished data). These data also highlight the importance of understanding the relative influence of multiple stressors on aquatic life degradation, because excess sedimentation at these sites may be a more immediate threat to stream biota at these sites, despite high nutrient concentrations. In arid environments like Utah, some streams are naturally turbid and may be less susceptible to the deleterious effects of nutrient enrichment. Stream metabolism metrics could provide a way of documenting that GPP and ER remain protective in highly turbid streams.

Rates of both GPP and ER are also influenced by natural conditions. For instance, sites with channel slopes < 1% had higher rates of GPP and ER than those above the threshold, although this relationship cannot be entirely attributed to natural conditions because slope was also strongly related to both TN (Pearson correlation r = -0.603) and TP (r = -0.617; data not presented in this TSD). The fact that nutrient concentrations were higher in lower gradient streams is not surprising considering that anthropogenic nutrient sources, including agricultural activities and urban discharges, are more likely to be concentrated at lower-gradient stream segments, where most of Utah's population resides. Residence times of nutrients are also lower in higher gradient streams, which could explain both the relationship with nutrients and also help to explain the influence of slope on stream metabolism via nutrient-related reductions in stream productivity. High flow events can also lead to much higher flow velocity among high-gradient streams. Such conditions could reduce GPP by scouring longer-lived, benthic macro algae (Grimm and Fisher 1989). ER rates could also be reduced due to exportation of accumulated organic matter (Acuna et al. 2004, Uehlinger et al. 2003).

Among the covariates evaluated, channel shading has the clearest influence on stream metabolism; the rates of both GPP and ER decline with increasing channel shading. Channel shading was only weakly correlated to TN (Pearson correlation r = -0.245) and TP (r = -0.221; data not presented in this TSD). Nutrients (TN and TP) and shading are both important determinants of stream metabolism, but the effects of these factors are very different and frequently unrelated. Nutrients elevate GPP and ER, whereas shading represses GPP by reducing the amount of light available for photosynthesis. As illustrated by the poor correlations, streams with high channel shading can occur in streams with high or low nutrients. The interactive effects of nutrient enrichment and channel shading on GPP rates are straightforward, and likely largely result from a reduction in light intensity and photosynthesis rates. The importance of light as a determinant in this relationship is supported by both experimental manipulations of stream shading (Kiffney et al. 2004) and landscape-level evaluations that assessed the relative importance of photosynthetic active radiation (PAR), nutrients, and other factors on GPP rates (Bernot et al. 2010, Bott et al 2006). The combined importance of light and nutrients to GPP rates has considerable management implications because it implies that potentially deleterious effects of excess nutrients can be minimized or exacerbated depending on the condition of the riparian corridor. This means that in many cases nutrient reduction efforts are more likely to improve biological conditions if they are accompanied with improvement in riparian conditions.

The relationship between lower ER values and streams with higher canopy cover is less intuitive than the corresponding reduction in GPP. One possible explanation is that GPP and ER are ecologically coupled processes, which is supported by the close correspondence between ER and GPP in landscape-level comparisons of GPP and ER in this and other investigations (Bernot et al. 2010). Further support is provided by observations that ER rates are tied to rates of bacterial (heterotophic) production (Suberkropp et al. 2010). The cause of the coupling of autotrophic and heterotrophic conditions remains unresolved, but one possibility is competition among these assemblages for important resources, like N and P. Some studies support the idea of competition as the driver of GPP and ER relationships; others do not. For instance, Carr and colleagues (2005) found a relationship between autotrophs and heteroptrophs in biofilm confirming the general association, yet they were unable to document competition among these assemblages for nutrients. It is possible that the relative importance of competition is dependent on ambient stream conditions, which is supported by observation that in high-nutrient streams algal-bacteria production become decoupled, whereas they remain closely coupled in low-nutrient environments (Scott et al. 2008). This idea is supported by Rier and Stephenson (2001) who found that algal cell biomass was the best predictor of bacterial cell density unless benthic chlorophyll-*a* concentrationwas low, which primarily occurred in low nutrient streams. From the perspective of resource management, the cause is probably less important than the general acknowledgement that the observation that ER, nutrients and channel shading interact to determine the relative sensitivity of streams to nutrient enrichment. Similar to GPP, the success of nutrient reduction efforts aimed at reducing ER rate is likely to be dependent on the health of the riparian ecosystem.

Summary and Recommendations

This study quantified the relationships between nutrients and stream metabolism (GPP and ER). Thresholds for GPP and ER were identified to quantify several condition classes based on these metabolism metrics. Perhaps most importantly, the comparisons of metabolism to DO criteria demonstrate that sites with high ER rates are also more likely to pose a threat to aquatic life uses. This correlation, coupled with the associated nutrient thresholds, will allow DWQ to more accurately identify sites with potential DO problems. If DO impairments are identified, GPP and ER data will also provide insight into the nature of the impairment. For example, if excessively high ER is observed at a stream with moderate or low GPP it would suggest that excessive carbon might be a more immediate nutrient-related concern than algal production and that some of the carbon sources likely originate outside the stream channel. Alternatively, the converse might be observed—high GPP, but low ER—without a DO problem during summer months, which may prompt the collection of additional DO data in autumn months as plants and algae begin to senesce.

The importance of covariates, especially slope and channel shading, was also highlighted. Excessive nutrients consistently played a role in the creation of high metabolic rates, but physical characteristics alter the relative susceptibility of steams to nutrient enrichment. DWQ will continue to apply metabolism data to other streams on a site-specific basis and will consider and quantify the influence of covariates on stream metabolism processes. This is particularly important for site-specific investigations, which provide the best opportunity to explore local physicochemical controls on GPP and ER rates. Once these controls are understood, DWQ can modify the regional thresholds to establish appropriate site-specific GPP or ER water quality goals.

Chapter 6

ORGANIC MATTER STANDING STOCKS

Key Points

Relationships between standing stocks of autochthonous organic matter (OM) and ambient nutrient concentrations were explored as a potentially temporally integrative response to nutrient enrichment.

Reach-scale estimates of autochthonous OM standing stocks in reference streams have a significant positive correlation to ambient total nitrogen (TN) and total phosphorus (TP) concentrations.

TN thresholds were identified that best differentiate streams with relatively low (< 0.238 mg/L), moderate, and high (> 1.95 mg/L) OM pools.

TP thresholds were identified that best differentiate streams with relatively low (TP < 0.026 mg/L), moderate, and high (TP > 0.589 mg/L) OM pools.

Once OM exceeds 48.76 g of ash free dry mass/m^2 (an OM standing stock associated with the high OM groups) the likelihood that dissolved oxygen falls below water quality standards increases considerably.

Introduction

Biogeochemical links among nitrogen (N), phosphorus (P), and carbon cycles in aquatic ecosystems have a rich history in ecological investigations (Redfield 1958). Anthropogenic increases in inorganic nutrients to streams have altered rates of accumulation, processing, storage, and transport of organic matter at local to global scales (Webster and Meyer 1997, Webster et al. 1990). At large scales, alterations to organic matter dynamics affect atmospheric and oceanic carbon cycles (Kominoski and Rosemond 2012). At smaller scales, such as stream reaches, changes in storage and transport of organic matter impact food resources (Hall and Meyer 1998, Hall et al. 2000), habitat availability (Walther and Whiles 2011, Yamamuro and Lamberti 2007), and ecosystem functions (Bilby and Likens 1980, Findlay et al. 2003).

Stream ecologists have focused on the role of organic matter budgets (Benstead et al. 2009, Fisher and Likens 1973), particularly those components that provide the energy base for stream food webs (Bonin et al. 2000, as reviewed in Tank et al. 2010). Early studies in organic matter budgets revealed that allochthonous organic matter (such as leaf litter) were the most important energy source in forested headwater streams (Fisher and Likens 1973). Other studies demonstrated that autochthonous energy sources become more important to food webs in larger, open-canopy streams (Hall et al. 2000, Minshall 1978). Recent research has shown that organic matter derived from algae is more readily consumed by stream microbes than organic matter derived from terrestrial sources (Lane et al. 2012, Ylla et al. 2012), which highlights the importance of understanding specific carbon sources.

Dissolved nutrient concentrations are known to stimulate organic matter processing rates in streams (Robinson and Gessner 2000, Triska and Sedell 1976). The relatively high N and P content of heterotrophic bacteria and fungi compared to the N and P content of particulate organic matter suggests that increases in dissolved nutrients will stimulate microbial activity (Stelzer et al. 2003). In fact, a nutrient-mediated increase in microbial—primarily fungal—biomass has been observed following several experimental nutrient additions (Fereira et al. 2006, Grattan and Suberkropp 2001, Gulis and Suberkropp 2002, Suberkropp et al. 2010) and respiration rates (Tank and Webster 1998, Young et al. 2008). These microbially mediated processes are a critical component of stream food webs because they convert more recalcitrant carbon, such as coarse particulate organic matter (CPOM), into more labile sources. Moreover, the microbes (i.e., bacteria and fungi) that colonize organic matter are a critical source of protein for macroinvertebrate shredders (Cummins et al. 1973) and may be the principal way in which allochthonous carbon enters detrital food webs (France 2011).

Increases in heterotrophic productivity can result from several anthropogenic drivers, including the quantity and quality of external (allochthonous) organic matter inputs to the system (Stelzer at al. 2003), primary production and associated autochthonous organic matter stream inputs, or increases in the rate of organic matter processing resulting from inorganic nutrient inputs (Tank et al. 2010). Regardless of the source, increases in heterotrophic productivity have implications for stream oxygen dynamics. Oxygen dynamics in streams are controlled by physical and biogeochemical processes. Daily changes in stream dissolved oxygen (DO) concentration primarily result from the biological processes of gross primary production (GPP) and ecosystem respiration (ER) and the physical process of reaeration (the exchange of gas between the stream and atmosphere):

$$\Delta DO = GPP - CR \pm reaeration$$

While decomposition of organic matter is a normal process in healthy streams, excess organic matter sometimes contributes to nighttime hypoxia and, less commonly, to anoxia, with deleterious effects to stream biota (Connelly et al. 2004, Kemp and Dodds 2001). The processes of oxygen consumption are often measured by the amount of oxygen consumed from the water column (biological oxygen demand [BOD]) or from the benthos (sediment oxygen demand [SOD]), or with reach-scale measures of stream metabolism (Acuna et al. 2004, Mulholland et al. 2001; see Chapter 5 of this technical support document [TSD]).

Rates of heterotrophic oxygen consumption are determined by the availability of nutrients (N and P) and by the availability and accessibility of organic matter (i.e., carbon). Therefore, organic matter standing stocks can be thought of as the potential for high rates of ecosystem respiration. Low minimum DO concentrations are offset by temporal site-specific factors (i.e., high reaeration related to flow or seasonally high GPP), which complicates the inference of potential problems with low DO from organic matter standing stocks. The risk of low DO problems certainly increases with increasing organic matter. In addition, exceptionally high organic matter alters stream food webs, which can potentially degrade biological uses. As a result, measures of organic matter standing stocks have promise as functional indicators of stream condition.

In this study the relationship between water column nutrient concentrations (N and P) and organic matter standing stocks in stream ecosystems was investigated to determine whether organic matter standing stocks increase in response to nutrient enrichment. The relationship between organic matter standing stocks and in-stream oxygen dynamics, including minimum DO concentrations and ecosystem respiration, was also investigated. Significant thresholds for each of the above relationships were derived to develop indicators of the effects of nutrient enrichment on organic matter standing stocks that contribute to DO criterion violations.

Methods

Field Methods

Organic matter standing stocks were surveyed at 35 stream reaches (Figure 3.1 and Appendix A) between September and November in 2010. At each site quantitative organic matter standing stocks were collected from a stream reach with a minimum length of 50 m using a 100-count, point-intersect method (after Bowden et al. 2006). Researchers walked from bank-to-bank upstream at a 45-degree angle relative to flow direction and sampled at five evenly spaced locations. At each point the channel unit type (rapid, riffle, glide, or pool) and the substrate type (fine benthic organic matter [FBOM], gravel/sand, cobble, wood, macrophyte, or filamentous algae) were recorded. Two to seven replicate samples (reps) were collected for each channel unit–substrate combination (store) sampled; the number of reps collected depended on the relative abundance of each store in the reach, as follows: < 10% = 2 reps, 10–39% = 3 reps, 40–59% = 4 reps, 60–79% = 5 reps, and 80–100% = 7 reps. Organic matter samples were collected using a stovepipe corer (FBOM, macrophytes, and filamentous algae), syringe (sand/gravel), or scraping biofilm from a specific area on hard substrate (cobble and wood). CPOM was isolated from other stores with a 1 mm sieve. Organic matter samples were placed in Whirl-pak® bags, placed on ice, and then filtered (FBOM, gravel/sand, cobble, and wood only) on pre-ashed Whatman GF/F filters and frozen within 16 hours of collection.

Surface water nutrients were collected on a minimum of three distinct sampling dates at the upstream and downstream ends of the reach between July and October and were analyzed at the Aquatic Biogeochemistry Laboratory at Utah State University (Valderrama 1981). Samples across time and location (a minimum of six) were averaged to determine water column nutrient concentrations for the reach. At each site a water quality probe (YSI 6600V2 or 600 OMS V2) was deployed to measure DO at 5-minute intervals for a minimum of 48 hours at the downstream end of each reach.

Laboratory Methods

Organic matter samples were subsequently quantified as ash free dry mass (AFDM) following established laboratory methods (Kiry et al. 1999). Samples were dried at 60^{o} C for 48 hours before being weighed on an analytical balance (Denver M-220, to 0.0001 g). Samples were then combusted at 450^{o} C for 2 hours, re-wetted, dried, and reweighed. AFDM was computed as the difference between the mass prior to ashing (organic plus inorganic matter) and subsequent to ashing (inorganic matter only). Larger samples of macrophytes and filamentous algae were subsampled prior to combustion and then scaled up to the entire sample. Multiple samples for each channel unit—substrate type were averaged. Organic matter areal concentrations were multiplied by the relative abundance of each channel unit–substrate type combination at the stream reach and then recorded as relative abundance per square meter.

Analytical Methods

Pooling Reach-Scale Data for Regional Analyses

Collection methods for organic matter (OM) standing stocks were fairly detailed—differentiating among several stream habitats and OM pools. These methods provide fairly accurate reach-scale estimates of OM standing stocks because they incorporate among-stream differences in OM sinks (they are scaled from the spatial extent of pools and riffles) and sources (different standing stocks are collected separately within each habitat). Such procedures will provide insight into the relative importance of OM sources and sinks at local scales. These analyses primarily aim to establish more general regional relationships between nutrients and OM standing stocks. As a result, the laboratory results of reach-scale OM stores were combined in a couple of ways. The first focus of the investigation was on OM stores that are most closely coupled to autochthonous production, including filamentous algae and the biofilm samples obtained from wood (epixylon), rock/hard surfaces (epilithon), and sand (episammon) (Table 6.1). FBOM was also investigated. FBOM could represent autochthonous sources or processed OM from allochthonous sources because these stores are important determinants of microbial and fungal heterotrophic production, which potentially can be directly influenced by inorganic nutrients.

Table 6.1. Relative amount of organic matter collected from each study reach, broken down by different stores. Mean nutrients are also provided for purposes of comparison. Site names and descriptions can be found in Table 3.1.

Site	Total Mass	FBOM g AFDM/m^2		Epixylon g AFDM/m^2		Epilithon g AFDM/m^2		Episalmon g AFDM/m^2		Fil. Algae g AFDM/m^2		Nutrients mg/L	
		Subtotal	%	Subtotal	%	Subtotal	%	Subtotal	%	Subtotal	%	TN	TP
BEC-AB	41.02	34.1	83.1	3.7	8.9	0.0	0.0	0.0	0.0	3.3	8.0	0.404	0.058
BEC-BL	256.00	252.0	98.5	3.2	1.3	0.0	0.0	0.0	0.0	0.7	0.3	6.351	0.838
BLACKFK	15.33	1.9	12.4	0.0	0.0	7.3	47.9	4.1	26.8	2.0	12.9	0.188	0.008
DCSP-AB	132.00	128.7	97.5	1.9	1.5	0.0	0.0	0.0	0.0	1.3	1.0	2.216	0.135
DCSP-BL	118.40	118.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.286	1.894
DIAFK	17.86	1.5	8.6	0.0	0.0	9.3	52.0	7.0	39.4	0.0	0.0	0.410	0.084
FISHCK	38.81	20.0	51.5	0.0	0.0	9.4	24.2	8.0	20.6	1.4	3.6	0.246	0.063
HUNTCK	23.76	1.2	5.0	0.0	0.0	22.2	93.3	0.0	0.0	0.4	1.8	0.455	0.076
KIMBALL	68.51	63.1	92.1	0.0	0.0	5.3	7.7	0.2	0.3	0.0	0.0	0.279	0.028
LBRAVON	27.00	5.0	18.7	0.0	0.0	19.0	70.4	2.9	10.9	0.0	0.0	0.343	0.021
LBRW-AB	15.95	12.5	78.3	0.0	0.0	2.9	18.1	0.6	3.7	0.0	0.0	1.175	0.075
LBRW-BL	7.79	3.8	48.6	1.1	14.2	1.8	22.8	1.0	13.4	0.1	1.0	1.085	0.084
LOGR1000	22.81	4.2	18.3	0.0	0.0	1.2	5.4	17.3	76.0	0.1	0.3	0.139	0.012
LOGRDUG	10.27	0.3	2.6	0.0	0.0	8.8	85.2	1.3	12.2	0.0	0.0	0.424	0.023
LOGRTB	5.38	1.6	29.5	0.0	0.0	1.7	30.9	0.2	3.7	1.9	35.9	0.132	0.012
MRTRE-AB	20.42	15.6	76.5	1.7	8.5	2.0	9.7	1.1	5.3	0.0	0.0	2.828	0.236
MRTRE-BL	44.90	40.2	89.5	3.8	8.4	0.5	1.0	0.5	1.0	0.0	0.0	3.897	0.445
NFCHLK	8.83	0.2	1.9	0.0	0.0	8.4	95.3	0.2	2.8	0.0	0.0	0.161	0.006
PRICER	13.80	4.5	32.7	0.0	0.0	0.0	0.0	9.0	65.4	0.3	1.9	0.388	0.050
PRP-AB	19.74	13.2	67.1	6.4	32.3	0.0	0.0	0.1	0.6	0.0	0.0	0.710	0.289
PRP-BL	37.96	37.1	97.8	0.2	0.6	0.5	1.4	0.1	0.3	0.0	0.0	2.950	0.732

Site	Total Mass	FBOM g AFDM/m^2		Epixylon g AFDM/m^2		Epilithon g AFDM/m^2		Episalmon g AFDM/m^2		Fil. Algae g AFDM/m^2		Nutrients mg/L	
		Subtotal	%	Subtotal	%	Subtotal	%	Subtotal	%	Subtotal	%	TN	TP
SALTCK	5.57	1.0	17.9	0.0	0.0	3.7	65.7	0.2	4.4	0.7	12.0	0.154	0.016
SCSNYD-AB	4.24	4.0	94.5	0.0	0.0	0.0	0.0	0.2	5.5	0.0	0.0	0.319	0.015
SCSNYD-BL	52.62	1.9	3.7	0.0	0.0	0.0	0.0	0.4	0.7	50.3	95.7	14.717	2.212
SFKLBR	13.54	10.8	79.4	0.0	0.0	1.9	13.8	0.9	6.8	0.0	0.0	0.229	0.017
SPRFV-AB	28.07	14.8	52.7	0.0	0.0	6.1	21.7	1.1	3.8	6.1	21.8	1.324	0.019
SPRFV-BL	30.68	11.1	36.3	0.0	0.0	10.2	33.2	8.6	27.9	0.8	2.5	1.677	0.078
SPRM-AB	60.44	13.8	22.8	0.0	0.0	1.1	1.8	10.0	16.5	35.6	59.0	1.232	0.033
SPRM-BL	194.48	165.7	85.2	0.0	0.0	0.0	0.0	28.7	14.8	0.0	0.0	10.416	7.897
TIEFK	37.49	17.1	45.5	0.0	0.0	0.1	0.1	18.6	49.7	1.8	4.7	0.118	0.007
UKMURD	34.08	5.2	15.2	0.0	0.0	23.6	69.1	5.0	14.7	0.3	0.9	0.109	0.004
UPRNFK	9.04	0.0	0.0	0.0	0.0	8.9	98.5	0.1	1.5	0.0	0.0	0.113	0.003
WEBR	43.64	0.6	1.3	0.0	0.0	41.5	95.1	1.5	3.3	0.1	0.3	0.382	0.040
WROAK-AB	19.03	1.4	7.3	0.0	0.0	12.0	63.2	3.8	20.1	1.8	9.4	0.109	0.009
WROAK-BL	15.01	0.4	2.6	0.0	0.0	5.7	38.3	1.7	11.1	7.2	48.1	0.125	0.020

Note: AFDM = ash free dry mass, TN = total nitrogen, and TP = total phosphorus.

Equally important were decisions about what OM standing stocks to exclude; contributions from macrophytes or coarse benthic organic matter (CBOM) were not included in total OM estimates. The presence and abundance of stream macrophytes is strongly influenced by physical factors such as flow regimes or substrate size (Chambers et al. 1991). In this study, macrophytes were entirely absent or in very low abundance at ~90% of the streams. Three of the four exceptions were located within the East Canyon Creek drainage where the wastewater treatment plant already limits nutrients to meet total maximum daily load (TMDL) expectations. Some macrophytes have the ability to assimilate nutrients from the water column and from sediments, and the relative contribution of each can change depending on site-specific nutrient availability and macrophyte species (Barko and Smart 1981, Madsen and Cedergreen 2002). As a result of these factors, macrophyte abundance at sites in this study was not significantly associated with nutrients (Table 6.2). Therefore, including macrophytes in these regional analyses would add unnecessary noise to the OM-water column nutrient relationships.

Table 6.2. Spearman rank correlation coefficients between total nitrogen and total phosphorus and each of the organic matter storage compartments evaluated in this study.

		Spearman Rank Correlation Matrix						
Organic Matter Compartment		FBOM	CBOM	Epixylon	Epilython	Episammon	Macrophytes	Fil. Algae
Total Nitrogen	r =	***0.56***	0.06	***0.45***	***-0.47***	-0.24	0.14	-0.14
	p =	***<0.001***	0.730	***0.007***	***0.005***	0.157	0.409	0.417
Total Phosphorus	r =	***0.56***	0.23	***0.52***	***-0.45***	-0.24	0.05	-0.17
	p =	***0.005***	0.185	***0.001***	***0.002***	0.173	0.786	0.335

Notes: Statistically significant correlations are in bold, italic text. CBOM = coarse benthic organic matter, FBOM = fine benthic organic matter.

CBOM was also excluded because it is made up of mostly OM that is terrestrial in origin (small branches, leaves, etc.); CBOM is a more important source of energy for invertebrates than heterotrophic microbes and fungi, at least until it is processed (Cummins et al. 1973). Moreover, the relative importance of invertebrates in OM processing decreases from upstream to downstream, which confounds measures of the relationships between nutrients and OM standing stocks.

These investigations were conducted in the summer and, therefore, missed autumn inputs, which is the most important period of heterotrophic inputs for temperate streams. CPOM constituted less than 6% of the OM standing stocks in late summer at the study streams. As a result, no significant relationships between nutrients and CPOM were found among these study streams (Table 6.2). As with macrophytes, the decision to exclude these pools from regional analyses does not suggest that these sources of OM are unimportant—they are; however, in this case sufficient information was not available to evaluate the components for purposes of the stressor-response analyses.

To obtain reach-scale abundance estimates, the remaining OM stores of interest were pooled based on their relative abundance along the 100 m study reach. First, within each reach, the relative abundance of each OM pool was calculated within riffles and then within pools. Next, to generate reach-scale abundance estimates, these data were scaled according to the spatial extent of the pools and riffles across the length of the reach. Finally, these reach-scale estimates were compiled to obtain the abundance of OM stores of interest for each stream.

Relating Stream Nutrients to Organic Matter

For these key OM stores, the extent to which OM standing stocks were correlated with either TN or TP concentrations was evaluated to help directly link the OM investigations to other indicators. Simple linear regression was used on log(x) transformed data to identify any linear relationships between nutrients and OM standing stocks. Nonparametric deviance reduction was used to identify TP and TN thresholds that separated OM standing stocks into distinct groups with maximal within-group similarity in the mean and variance of chemistry samples (Qian et al. 2003, package rpart). Significance of multiple threshold models was tested using analysis of variance (ANOVA) followed by a Tukey's honestly significant difference (HSD) test.

Relating Oxygen Dynamics to Organic Matter

The relationship between OM standing stocks and in-stream oxygen dynamics was evaluated. A consistent association between increased OM standing stocks and violations of DO may indicate a causal link between increased OM standing stocks and the degradation of aquatic life uses.

DO measurements taken every 5 minutes over a 48–72-hour period were used to calculate the observed minimum DO (mg/L). With two exceptions, the aquatic life uses assigned to each site were used to evaluate the percent of samples that exceeded Division of Water Quality (DWQ) minimum and 30-day average DO criteria (Utah Administrative Code [UAC] R317-2-14; Table 6.3). Two sites were designated with habitat-limited uses (3D), which do not have DO criteria; DO criteria for 3C uses were used for these sites.

Table 6.3. Dissolved oxygen criteria for different aquatic life use designations applied to Utah streams.

Aquatic Life Use Designation	Minimum DO Criterion mg/L	30-day Average DO Criterion mg/L
Coldwater Fish (3A)	4.0	6.5
Warmwater Fish (3B)	3.0	5.5
Nongame Fish (3C)	3.0	5.0

Source: UAC 317-2-14, Table 2.14.2
Note: DO = dissolved oxygen.

Relating Organic Matter to Ecosystem Respiration

OM standing stocks were compared with ER (see Chapter 5), because excess carbon is potentially as important as TN and TP in causing low DO within streams. As with nutrients, nonparametric deviance reduction techniques were used to determine thresholds of OM standing stocks that best identify sites that violate DO criteria and ER values of potential concern. The significance of thresholds identified via these deviance reduction analyses were then tested with student's t-tests; Wilcoxon rank sum tests were used if statistical assumptions of parametric methods were violated.

Evaluating the Influence of Physical Covariates

OM sampling methods for this study focused on the storage of autochthonous OM (i.e., excluding CBOM), because these are the carbon stores that are directly influenced by nutrient inputs to streams. Even so, a large portion of the stored OM within streams may be derived from terrestrial sources and stored as FBOM or incorporated into biota. As a result, OM standing stocks will vary, not only with nutrients, but also with a number of physical factors unique to the watershed and riparian corridor. Therefore, the relative importance of nutrient concentrations and watershed characteristics on among-stream differences in OM standing stocks was evaluated.

Random forest regression models (Breiman 2001, package randomForest) were fitted to predict OM standing stocks from stream nutrients (TN and TP) and other chemical constituents, watershed characteristics derived from U.S. Geological Survey (USGS) Stream Stats (USGS 2014), and site-specific physical habitat measures from DWQ's Utah's Comprehensive Assessment of Stream Ecosystem protocols (DWQ 2012). The relative signal from nutrients and the most important physical covariates (determined from Random forest regression) were compared to separate nutrient-OM standing stock relationships from other potential sources of variation.

Random forest regressions were used to identify the suite of candidate covariables that best predict among-stream differences in OM standing stocks. Random forest regressions build multiple successive regression trees using "bagging," which roughly equates to more common bootstrapping procedures. Specifically, multiple regression trees are constructed, each one based on a subsample of observations and a subset of available data (Breiman 1996). In each case, regression trees are built with the best possible splits among the subset of predictor variables to obtain the most accurate regression tree possible. Each tree then gets a vote in the final prediction, which is based on the ensemble of trees—the forest. Once forests were constructed, the relative importance of potential predictor variables was determined using the variable importance (VarImp) procedure within the random forest package. The VarImp procedure randomly reassigns (shuffles) observations for each potential predictor one at a time before re-running the forest models. The increase in mean squared error (MSE) that results from the new Random forest regression quantifies the relative importance of that variable to the Random forest regression. This procedure is then repeated for each candidate variable, one at a time. Once complete, the relative importance of predictors can be determined. Those variables that lead to the greatest reduction in predictive accuracy once observations have been randomly assigned are assumed to be of greatest importance. Random forest models were chosen to determine variable importance because these approaches are unbiased by datasets of highly correlated variables, robust against overfitting (due to the randomization procedures), and do not require adherence to parametric statistical assumptions (i.e., homoscedasticity, normal distributions) (Breiman 2001).

It is important to explore potential bias among all covariates while seeking a model that is readily interpretable and parsimonious. As a result, a second random forest model was fitted with only the most important variables. The reasoning behind this step was that if a second model with fewer variables performed as well as the larger model, then it would be easier and more cost-effective to measure important covariates in future management applications. All analyses were conducted in R v2.15.0 (R Core Team 2012).

Results

General Patterns with Organic Matter Standing Stocks

OM standing stocks varied by thee orders of magnitude among all sites, with total stores ranging from 4.24 to 256.0 g AFDM/$m^{2,}$ and a median of 23.76 g AFDM/m^2. FBOM was the largest contributor to reach-wide OM, with an average of 45% of the total standing stock; FBOM was followed by epilithon (30.5%), episammon (13.1%), filamentous algae (9.2%), and epixylon (2.2%) (Table 6.1).

Relationship between Organic Matter and Nutrients

Among-steam nutrient concentrations were also quite variable with a log-normal distribution. TN varied from 0.109 to 14.72 mg/L, with a median of 0.404 mg/L. TP varied from 0.003 to 7.89 mg/L, with a median value of 0.04 mg/L. After log transformation, significant relationships between OM standing stocks and TN were identified using linear regression ($r^2 = 0.40$, $p < 0.001$) and TP ($r^{2=}$ 0.39, $p < 0.001$; Figure 6.1).

The extent to which different stores of OM were related to nutrients was evaluated; among all sites only FBOM, epixylon, and epilithon varied significantly with TN and TP (Table 6.2).

Identification of Nutrient Thresholds

Nonparametric deviance reduction was used to identify thresholds of TN and TP that best grouped OM into distinct groups. For TN, thresholds at 0.238 and 1.95 mg/L TN separated OM standing stocks into three groups (L = low, M = medium, and H = high) with statistically different N concentrations (ANOVA $p = < 0.001$, Tukey's HSD L-M $p = 0.26$, L-H $p < 0.001$, and M-H $p = 0.002$; Figure 6.2). Using the same procedure for TP, statistically significant thresholds at 0.026 and 0.589 mg/L were identified (ANOVA $p < 0.001$, Tukey's HSD L-M $p = 0.01$, L-H $p = < 0.001$, and M-H $p = 0.006$; Figure 6.2).

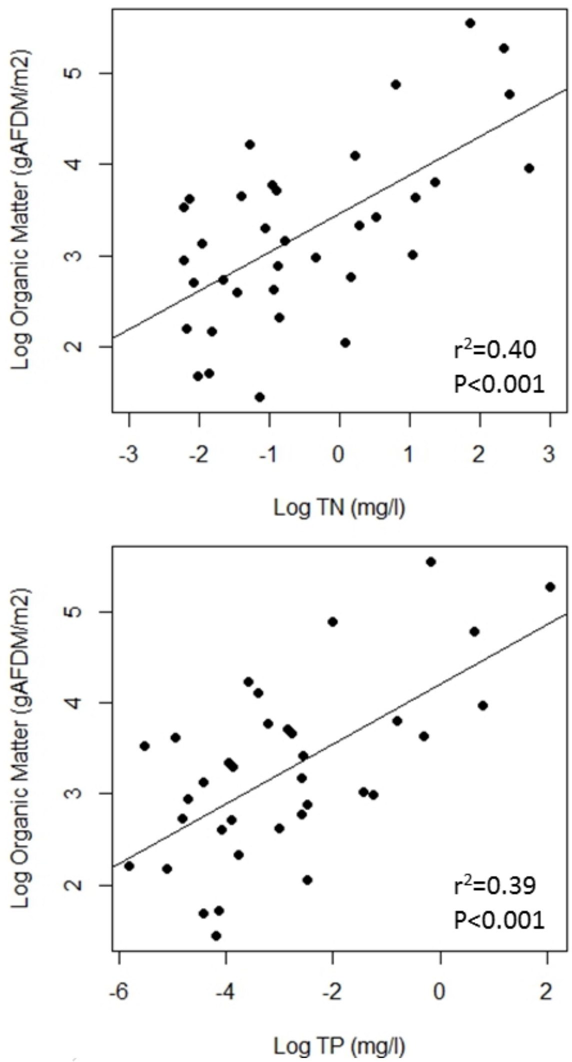
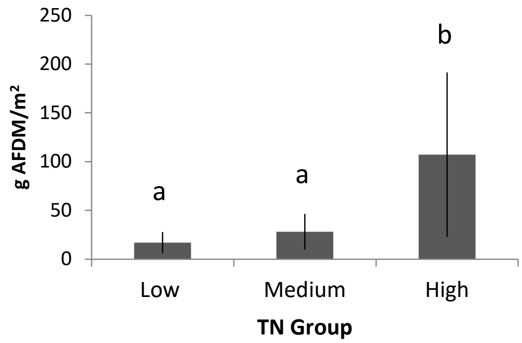


Figure 6.1 Linear regression between surface water organic matter and total nitrogen (TN, top panel, $r^2 = 0.40$, $p < 0.001$) and surface water organic matter and total phosphorus (TP, bottom panel, $r^2 = 0.39$, $p < 0.001$) and ash free dry mass (AFDM, g/m^2).



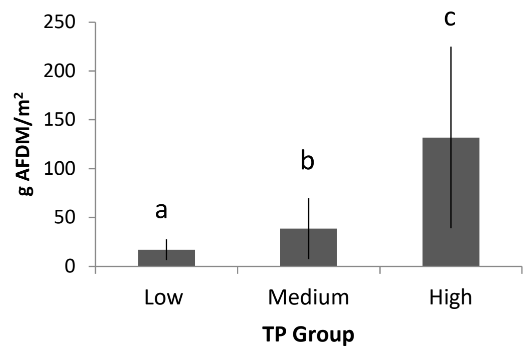


Figure 6.2. Organic matter standing stocks (ash free dry mass [AFDM] g/m^2) among streams within low, medium, and high nutrient groups. Bars with different lowercase letters above them are significantly different from one another (analysis of variance test and *post hoc* Tukey's honestly significant difference test); error bars are one standard deviation.

Relationships with Existing Dissolved Oxygen Criteria

OM standing stocks were strongly associated with DO concentrations obtained from 48–72-hour probe deployments. Minimum DO values among all sites varied from 0.39–8.53 mg/L. Overall, 31 of the 35 sites never fell below the minimum (acute) DO criterion, but in the 4 sites where excursions did occur DO values were below the criterion for approximately one-third of the day. DO fell below the 30-day average (chronic) DO criterion at 14 of the 35 stream sites. Among the 14 streams where DO fell below chronic thresholds, the amount of time the streams remained below the criterion varied considerably, from 3% to 75% of 5-minute interval samples. Daily ER rates were calculated for 31 sites (see Chapter 5 for details) and were significantly, albeit weakly, correlated to OM standing stocks (linear regression $r^2 = 0.15$, $p = 0.02$, data not presented in this TSD).

Determining whether sites above and below DO benchmarks could be distinguished with OM standing stocks was also an area of interest. Nonparametric deviance reduction was used to identify independent thresholds of OM standing stocks for each of the four DO benchmarks. In all four cases, a threshold of 48.76 g AFDM/m^2 was identified. Each of these thresholds divided streams into groups with statistically significant ($p < 0.05$) differences in OM standing stocks (Table 6.2, Figure 6.3)

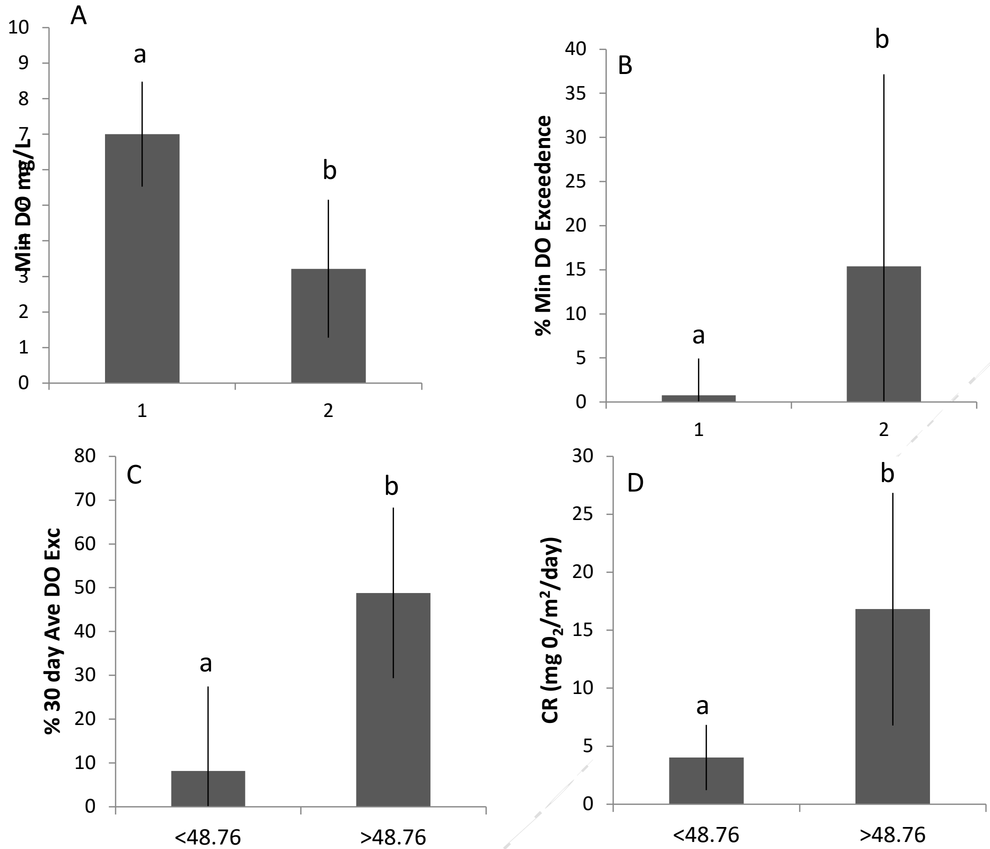


Figure 6.3. Relationships among several water quality benchmarks and among streams with low (< 48.76 g ash free dry mass/m^2) and high organic matter. Data for the water quality relationships were obtained from a minimum 48-hour sample period. Panel A depicts the minimum dissolved oxygen observation of this period. Panels B and C depict the number of 5-minute observations that fall below the two numeric dissolved oxygen criteria that differ with respect to their averaging periods. Panel D compares ecosystem metabolism among streams with low and high organic matter. Bars with different lowercase letters above them are significantly different from one another determined by student's t-tests (A and D) or Wilcoxon rank sum test (B and C). Error bars are one standard deviation.

The Influence of Physical Covariates

The random forest model based on all candidate variables was significant, albeit with fairly low accuracy (pseudo $r^2 = 0.239$). Nevertheless, the model did elucidate the relative importance of the variables evaluated. TN and TP scored high on measures of variable importance (68.4% and 53.8% increase MSE from VarImp procedure). The watershed characteristic percent fast water habitat (% riffles + % rapids) was the most important variable with an 87.4% increase in MSE. Two reach-scale physical habitat parameters, mean wetted width (m) and watershed area (miles^2), were about half as important as nutrients (36.1% and 26.0% increase MSE, respectively) in explaining among-stream OM differences. A second random forest model was fitted with only these five variables, and model performance increased (pseudo $r^2 = 0.431$). The final model ranked variable importance in identical order as the first model with the full suite of parameters. Percent fast channel (123.9% MSE) was most important followed by TN (97.0% MSE), TP (79.2% MSE), watershed area (42.9% MSE), and mean wetted width (24.0% MSE) (Table 6.4).

Table 6.4. Relative importance of ambient nutrients and other covariates in prediction of stream organic matter for one model that includes all possible variables and another limited to those variables with the most explanatory power.

Environmental Gradient	Units	Full Model % Increase MSE	Subset Model % Increase MSE	Source
Total Nitrogen*	mg/L	68.4	89.1	USU ABL
Total Phosphorus*	mg/L	53.8	81.7	USU ABL
Turbidity	NTU	15.5		UPHL
Total Suspended Solids	mg/L	20.6		UPHL
Slope	%	19.9		USGS
Basin Area+	mi^2	26.0	39.4	USGS
Herbaceous Upland+	%	5.8		USGS
Forested Watershed+	%	2.2		USGS
Basin Slope+	%	0		USGS
Mean Water Depth	cm	12.4		DWQ
Mean Wetted Width	m	36.1	14.4	DWQ
Mean Thalweg Depth	cm	13.6		DWQ
Bankfull Height	cm	-4.8		DWQ
Channel Incised Height	cm	1.8		DWQ
Channel Width:Depth	ratio	14.5		DWQ
Fast Water Habitat	%	87.4	137.9	DWQ
Riparian Corridor Bare Ground	%	-8.5		DWQ

Notes: Nutrients (*) are included for comparison purposes. Environmental gradients (covariates) include field measurements taken at each stream and watershed-scale attributes (+). Results were generated from random forest models that predict GPP and ER from all variables in the table. Mean squared error (MSE) is a measure of the increase in model error that results when each variable is randomized among sites, while keeping all other variable constant. Data sources include: Utah State University Aquatic Biogeochemistry Laboratory (USU ABL), Utah Unified Public Health Laboratories (UPHL), U.S. Geological Survey Stream Stats program (USGS) or the DWQ Comprehensive Assessment of Stream Ecosystems program

Discussion

Organic Matter Standing Stocks

A significant relationship among OM standing stocks and nutrient concentrations (TN and TP) was identified. OM standing stocks had a relatively strong linear relationships with both TN ($r^{2=}$0.40) and TP ($r^{2=}$0.39), especially considering the diversity of streams sampled. This confirms that the sampling methods chosen specifically to detect the portion of OM standing stocks that are influenced by in-stream nutrients (i.e., excluding CBOM) was successful.

Of all the storage compartments evaluated for OM standing stocks, FBOM was the largest. While FBOM is certainly an important sink of OM in streams, it may also produce some of the error in the relationships with in-stream nutrients or the ER responses. Given that these data were collected in autumn, some of the FBOM was undoubtedly from autochthonous sources. However, the FBOM also contained carbon of terrestrial origin that was broken down into smaller particles and stored as FBOM. Terrestrially derived FBOM would not be expected to be associated with stream nutrients except in the sense that high-nutrient streams may have lower allochthonous:autochthonous FBOM due to increases in carbon processing rates. The fact that ER was higher in high-carbon streams suggests that such higher rates of OM processing may be occurring within these systems. In this context, the relationships between nutrients and OM would be the inverse of autochthonous sources because labile OM sources are consumed by heterotrophs at a higher rate, which would decrease the strength of these relationships. Increases in carbon processing rates have been directly quantified by others. For instance, Benstead and colleagues (2009) showed that artificial N and P additions to detritus-fed streams increased processing rates of CBOM when compared to FBOM. This indicates that, although FBOM in the study sites may originate from terrestrial or aquatic sources, the high levels of FBOM may be indicative of responses to nutrient enrichment. Future OM research by DWQ will quantify the $^{13}C/^{12}C$ of FBOM, which may help better elucidate the ultimate source of carbon inputs (Palmer et al. 2001). N isotopes may also be useful because aquatic plants and algae are typically enriched by 3% $^{15}N/^{14}N$ compared to terrestrial counterparts (French 1995).

Nutrient Thresholds

Thresholds of TN and TP that best separated OM standing stocks into three groups of streams were identified. On average, a TN concentration of 0.238 mg/L distinguished between streams with low to moderate levels of OM, whereas a concentration of 1.95 mg/L distinguished between streams with moderate and high levels of OM. For TP the lower threshold was 0.026 mg/L, and 0.589 mg/L defined the upper threshold. Differences between OM standing stocks were less distinct between the low and medium nutrient groups than for the medium and high nutrient groups. In the case of N, OM standing stocks did not differ between TN-low and TN-medium groups (Figure 6.2). DWQ will use these thresholds, in conjunction with others, to identify sites with potential nutrient-related problems. However, these nutrient thresholds are among the weakest of investigated responses, perhaps because the origins of FBOM could not be definitively determined to be allochthonous or autochthonous.

Organic Matter and Dissolved Oxygen

If OM standing stocks will be used as an indicator of anthropogenic eutrophication, it is important that any thresholds identified are both statistically significant *and* representative of potentially deleterious effects on aquatic life uses. To examine the ecological relevance of these thresholds, OM standing stocks were compared to DO metrics (minimum DO and ER) and to Utah's DO water quality criteria (minimum DO standard and 30-day average DO standard).

The majority of sites (89%) did not show a violation of the minimum DO criterion during the study period. Among sites where DO fell below water quality benchmarks, periods of low DO were long, which makes it highly likely that DO threatens aquatic life at these locales. Importantly, these extreme circumstances would likely have been missed with routine grab samples, because standards were not exceeded for two-thirds of the day at times when grab samples are most likely to occur. Ideally, DO assessment decisions would always be made with quality, high-frequency data, yet these data are not always readily available for assessment purposes. Another important consideration when interpreting these DO results is that they only capture a 3–7-day snapshot of DO conditions. Periods of low DO may occur during other times of the year, particularly at sites where clearly living OM standing stocks are high, because ER would increase during autumn senescence as living OM becomes more labile; moreover, the ability of GPP to offset losses of DO from ER may also be diminished during these periods due to decreases in temperature and algae abundance A more temporally stable surrogate measure, like OM standing stocks, may help identify sites with potential DO problems.

Aquatic biota can also be negatively affected by chronic (long-term) exposure to low DO. Chronic effects of low DO were estimated by comparing the percent of times DO fell below the 30-day average minimum DO criterion and relating that data to OM standing stocks (Table 6.3). Samples at sites with organic matter > 48.76 g AFDM/m^2 were lower than their 30-day DO criterion 48.8% of the time, whereas this only occurred 8.8% of the time at sites with low OM standing stocks. These relationships do not necessarily imply impaired conditions because the data are temporally limited. Some sites that are fully attaining aquatic life uses may occasionally fall below the 30-day average criterion without harm to aquatic life uses. Nevertheless, these data provide credence to the use of OM as a potential screening tool for sites with potential DO problems, especially considering that OM thresholds were identical for other DO water quality benchmarks that were evaluated.

Organic Matter and Metabolism

As expected, sites that had greater OM standing stocks had higher rates of ER. In fact, ER OM thresholds were identical to those derived from DO water quality criteria. While identical OM thresholds were not expected among these indicators, it is not surprising that they are related. The four oxygen metrics were not completely independent as they were calculated from the same DO record. Eutrophication, by definition, increases autotrophic production and biomass in aquatic systems.

A less frequent, but potentially equally important consideration is that nutrient enrichment can also stimulate heterotrophic productivity and increases in OM processing rates (Robinson and Gessner 2000), which further reduces DO in streams. In the Benstead and colleagues (2009) study mentioned above, they found increased heterotrophic respiration rates per gram of substrate on leaf litter, woody debris, and FBOM after two years of whole stream nutrient additions. This study agrees with previous literature that has shown a positive effect of nutrient enrichment on microbial respiration (Cross et al. 2006, Gulis and Suberkropp 2002). The study reported here suggests that heterotrophic responses to nutrients may hold true across a broad range of streams with different OM composition and stocks.

Physical Covariates

Numerous stream physical characteristics affect the delivery, storage, and transport of OM within streams. The most import variables identified with Random forest regression models are consistent with OM processes. Among all the variables evaluated, the percent fastwater habitat (transport), nutrients (TN, TP, storage/processing), and stream size (watershed area and mean wetted width, delivery, and storage) were the strongest predictors of OM storage. These observations are consistent with accepted stream ecology theories such as the river continuum concept (RCC; Vannote et al. 1980) that predicts changes in physical and ecological characteristics from headwaters to the valleys. Increases in water velocity (likely due to increased slope) are all likely to occur at higher-order streams where the majority of OM inputs are from finely processed terrestrial matter (CBOM). All the important physical covariates quantify changes from smaller headwater streams to the valleys. The influence of covariates does not preclude the potential importance of TN and TP. The random forest models suggest that both of these macronutrients have a more important influence on OM standing stocks than stream size, which in turn suggests that biogeochemical processes, not just stream size, are important factors influencing the storage and retention of OM within streams.

Summary and Recommendations

Relationships were established between nutrients and OM and between OM and DO dynamics. Significant thresholds were developed for each of these relationships that can help guide future monitoring and assessment efforts, ultimately helping to identify potential solutions to streams with nutrient-related impairments. The study demonstrated that increased OM standing stocks are associated with lower minimum DO concentrations and more violations of DO criteria, which provides a link between OM indicators and support of aquatic life uses. The multivariate analyses show that among-stream differences in OM standing stocks could not be explained by physical characteristics alone, and that excess nutrients play a role in OM accumulation and storage. An OM standing stock threshold of > 48.76 g AFDM/m^2 is suggested as a broadly applicable regional indicator of nutrient enrichment.

Follow-up OM studies will be most useful as a way to guide site-specific investigations among streams where low DO concerns have been identified because high-frequency DO data are increasingly inexpensive and easy to collect. In cases where DO concerns have been identified, the OM thresholds could be used to infer whether OM pools are atypically high. Insights gleaned from these investigations will prove insightful, for instance, in TMDL studies addressing low DO problems. In these circumstances, OM standing stocks may be a useful water quality objective—provided that local conditions such as slope are accounted for—because it provides a time-integrative measure of an important driver of low DO. Also, in some circumstances reductions in OM standing stocks may be a direct measure of best management practices that may be employed to address DO concerns, which may make this indicator a robust indicator of incremental progress.

These data also highlight the importance of confirming these indicators on a site-specific basis, which will remain an integral part of Utah's nutrient reduction efforts. Additional efforts will need to further elucidate the relative roles of nutrients and important covariates. Specifically, follow-up investigations will need to address the site-specific importance of habitat characteristics associated with OM transport (i.e., slope and habitat complexity). These site-specific investigations should also identify important OM sources. Additional details provided by these site-specific investigations will more accurately characterize the causes of DO impairments and the influence of increased nutrients on stream food webs. Such insights, together with other lines of evidence discussed in this report, should be incorporated into future remediation efforts.

Chapter 7

STRUCTURAL INDICATORS: EFFECTS OF NUTRIENT ENRICHMENT ON THE COMPOSITION OF STREAM BIOTA

Key Points

Historical data from DWQ's biological assessment program were used to explore changes in biological composition as a structural indicator of nutrient enrichment.

The greatest changes in the composition of diatoms were associated with ambient total phosphorus (TP) of 0.022 mg/L (nCPA, TITAN), with the most sensitive taxa exhibiting the largest response at 0.016 mg/L and the most tolerant taxa at 0.042 mg/L.

Compositional changes to macroinvertebrates with primarily explored using O/E because this index is routinely used by DWQ to evaluate support of aquatic life uses.

Significantly significant declines in O/E were observed as both ambient TP and total nitrogen (TN) increased among study streams.

A TP of 0.045 mg/L and TN of 0.41 mg/L best distinguished between sites considered to be impaired by O/E assessments as opposed to sites considered to be fully supporting of aquatic life uses.

Introduction

Worldwide, aquatic resource managers use numerous measures of stream condition to identify water quality problems. Bioassessments—quantitative descriptions of anthropogenic alterations to the composition or structure of aquatic assemblages—are among the most meaningful assessment tools. Resident aquatic communities integrate the effects of stressors through time because they are subjected to long-term impacts to ecosystems over weeks or years. Biological assessments also integrate the effects of multiple stressors, both spatially and temporally (Fausch et al. 1990). They are also of direct interest to the public, and this interest is expressed in the goals of many regulations that seek protection of aquatic resources. In the United States, the support and maintenance of biological integrity is one of the fundamental objectives of the Clean Water Act (CWA 101(a)).

Excess nutrients are one of the greatest threats to the biological integrity of the nation's waters (U.S. Environmental Protection Agency [USEPA] 2000). Anthropogenic sources of nutrients to the nation's waterways have been a known stressor to aquatic communities prior to the adoption of the CWA (e.g., Carr 1962, Vallentyne 1974). Aquatic biological communities, especially diatoms and macroinvertebrates, are particularly sensitive to excessive surface water nutrients (King and Richardson 2003, Smith et al. 2007, Van Sickle and Paulsen 2008, Wang et al. 2007). At very high concentrations, nutrients—particularly nitrate and ammonia—can be acutely toxic to biota, but such conditions rarely occur in streams. More often, human nutrient inputs to streams alter ecosystem processes (i.e., dissolved oxygen fluctuations, organic matter [OM] processing), which in turn affects the structure and condition of biotic food webs. Structural

Biological Integrity: An Objective of the Clean Water Act

The Clean Water Act seeks to protect, maintain, and restore the biological integrity of the nation's waters. This objective is built upon the concept that increasing human activity alters stream function and structure. Currently the generally accepted definition of biological integrity (after Frey 1977), is "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region." No single measure can capture all the ecological attributes captured in this definition, but various indicators have been established to estimate the degree of departure based on alterations to the composition of stream biota.

indicators of nutrient enrichment provide quantitative estimates of the extent of alteration to biological assemblages.

Not all species respond similarly to increased nutrients; some species cannot tolerate changes caused by excessive nutrients, whereas others are adapted to such conditions and thrive in nutrient-enriched conditions (Davies and Jackson 2006). Because of these differences, it is important that metrics used to identify structural responses to nutrient enrichment account for orthogonal responses among taxa. Structural responses also vary among sites with different environmental conditions (i.e., channel shading, temperature, and substrate characteristics), and bioassessments need to account for the influence of these covariates to accurately estimate the composition expected under unaltered conditions.

Different assemblages (i.e., diatoms versus macroinvertebrates) can also differ with regard to their relative sensitivity to nutrient enrichment. Some assemblages, like algae and diatoms, respond directly to nutrients; other assemblages respond indirectly because other changes must occur before they are affected. For instance, as nutrients at a site increase, diatoms are often replaced by other algae taxa; this replacement can then affect invertebrates—either positively or negatively—due to the resulting changes in habitat or food quality (i.e., diatoms to filamentous algae; Peterson et al. 1993). Compositional changes are not independent, so it is not necessarily appropriate to give deference to one assemblage over another. Instead, regional thresholds are best derived from the weight of evidence obtained from several different measures of condition. Ideally, they would reflect several aspects of food web dynamics (i.e., different trophic levels, functional feeding groups).

Measures of biological integrity involve multivariate comparisons of compositional similarity. Using several analytical approaches when developing quantitative biological assessment tools leads to a more robust analysis than relying on a single approach. Convergence among thresholds obtained from different analytical techniques improves confidence that thresholds are environmentally meaningful and not simply statistical artifacts.

The primary objective of this chapter is to report on N and P concentrations found to best explain among-stream differences in the composition of diatom and macroinvertebrate assemblages. Different taxa (i.e., genera or species) were evaluated using several analytical approaches to see whether the different taxa responded systematically to varying nutrient concentrations. Individual diatom and macroinvertebrate taxa that consistently increase or decrease in abundance among streams with varying nutrient concentrations were analyzed. To identify these taxa, the threshold indicator taxon analysis (TITAN; Baker and King 2010, King and Baker 2010) was used, which is a recently proposed analytical approach derived from long-established ecological relationships. TITAN integrates bidirectional taxa occurrences (presence versus absence) and relative abundance in relation to an environmental stressor to determine stressor-response thresholds. TITAN identifies thresholds for individual taxa and integrates them into single thresholds to establish a regional response threshold for each assemblage. These taxa-specific and integrative thresholds are established for taxa that respond positively (increase in abundance) and negatively (decrease in abundance) to increasing nutrients. This makes TITAN an ideal technique to derive structural thresholds because it provides stressor 'bookends' with one threshold that identifies concentrations where the most sensitive taxa are lost and another that identifies conditions where tolerant taxa thrive (Kail et al. 2012, King et al. 2011). TITAN measures changes in composition, but such compositional changes do not necessary imply degradation of aquatic life uses.

Another approach to threshold derivation evaluates the nutrient concentrations that most closely correspond to independently derived measures of aquatic life degradation. The Utah Division of Water Quality (DWQ) calculates macroinvertebrate observed/expected (O/E) ratios for bioassessments (DWQ 2016 Integrated Report). These estimates of condition are derived from analytical techniques that have been employed in numerous settings for over two decades. In brief, O/E ratios are derived from empirical models that compare the taxa expected (E) at a site without anthropogenic degradation to the predicted taxa that are actually observed (O). The E values are derived from models that use compositional differences among regional reference sites to make site-specific predictions of expected taxa. To make these predictions, the models use site-specific measures of natural environmental gradients—those unlikely to respond to human-caused stressors (geology, geography, precipitation, etc.)—to predict the probability of capturing each of the taxa (i.e., genus or species) that are part of the regional species pool. Once created, these models can be used to generate biological expectations for other sites (i.e., those not used in model development) based on their site-specific physical and geographical characteristics. The biological expectations, expressed as probability of capture, are provided for all known taxa within the modeled region (Hawkins et al. 2000).

DWQ recognizes that other methods of biological condition assessment are available and that the two models described above are not exhaustive in their ability to account for site-specific covariables. However, the models do account for broad physical and geographical variables and, because they are used by DWQ as quantitative estimates of the support of aquatic life uses, O/E can be used to estimate N or P thresholds that, on average, are associated with independently derived measures of biological degradation.

This chapter uses existing biological assessment data to answer several important questions related to the derivation of nutrient criteria for Utah's streams:

1. What concentrations of N and P are associated with the largest changes in the distribution and abundance of sensitive macroinvertebrates and diatom taxa? How do these concentrations differ for tolerant taxa?
2. What concentrations of N and P best distinguish between biologically degraded and nondegraded streams? Do these thresholds differ based on the analytical method used?
3. Do sites that exceed N or P thresholds correspond with sites that would be considered biologically degraded based on independent biological assessments?

Methods

Stream Sites (Data Selection)

The diatom and macroinvertebrate data used for these analyses were collected between 2001 and 2010 from 370 streams across Utah. Several sources were used, including Utah's Comprehensive Assessment of Stream Ecosystems, the National Wadeable Streams Assessment (WSA), National Rivers and Streams Assessment (NRSA), DWQ specific programmatic sampling events (standards development), and the U.S. Geological Survey's National Water-Quality Assessment (NAWQA) Program (Cuffney et al. 1993). Field methods among programs followed nearly identical sampling procedures (USEPA 2007). Irrespective of the data source, water chemistry and habitat characteristics were collected immediately prior to the collection of biological samples.

While protocols were nearly identical among collection programs, several differences in laboratory methods affected how the data were treated. One important difference was whether or not total N (TN) data were available. TN was only recently added to DWQ's regular chemical analytical suite, which limited the number of sites DWQ was able to evaluate and report on in this chapter. Laboratories also differed with respect to their reporting limits for different parameters. Many samples came from DWQ studies, and the Utah State Health Laboratory where the samples were processed has a detection limit of 0.02 mg/L total P (TP), while other samples used in these analyses (WSA, NRSA, and NAWQA) were analyzed at laboratories that had lower detection limits. In all cases, a value of one-half of the lab-specific reporting limit for was used for subsequent analyses in this investigation.

Biological Data Collections

Biological samples were obtained following USEPA's rapid bioassessment protocols standard operating procedures (Barbour et al. 1999). DWQ's diatom bioassessment program is relatively new, so although all sites with diatom data had corresponding macroinvertebrate data, the converse was not always true. This limitation, coupled with limited TN data (see above), means that the number of sites differed for each analysis because the number of sites was maximized for each different type of analysis.

Diatoms were collected from hard surface benthos—typically cobbles—from 11 transects at each stream. Diatom samples were chilled (for preservation) and then subsequently identified to lowest practical taxonomic resolution by Rushforth Phycology, LLC following standard laboratory procedures (Rushforth Phycology 2005). Macroinvertebrate data are based on a composite of 8 fixed-area riffle samples. Macroinvertebrate samples were preserved in ethanol and processed to lowest practical taxonomic resolution by the Bureau of Land Management BugLab using 500-count fixed-count subsample methods (Miller and Judson 2011).

Analytical Methods

Compositional Changes (TITAN)

TITAN in R v2.15.0 (R Core Team 2012) was used from package TITAN, and the analytical procedures described in Baker and King (2010) were followed. In all, 370 sites in Utah had both diatom and macroinvertebrate data. Of the available sites, there were too few with both diatom and TN data to create TITAN models, so TITAN diatom thresholds were not generated for TN. Of the 370 sites, 251 had both diatoms and TP data, which was more than sufficient for calculating TP thresholds from TITAN. Of the 370 sites, 178 had TP, TN, and macroinvertebrate composition data, so subsequent TITAN models were limited to these sites for this assemblage to be consistent in evaluation of TN and TP stressors. For both assemblages, TN and TP were used independently as environmental stressors in TITAN analyses.

In this analysis, an exhaustive list of site-specific covariables was not incorporated into the TITAN analyses. Site-specific covariables could be assessed, as appropriate, in development of site-specific criteria (See Chapter 12).

Biological Impairments

River invertebrate prediction and classification system (RIVPACS) models created for DWQ's 2010 *Integrated Report* were used to generate O/E scores (DWQ 2010, see pages 31–36 for details). Macroinvertebrate O/E scores were calculated with data collected from 243 stream sites. These sites include 97 reference sites and 146 randomly selected sites that best represented the range of conditions found throughout the state (Olsen and Peck 2008). All 243 sites were used to make O/E and TP comparisons; TN comparisons were limited to only 68 sites for which both TN concentration data and O/E scores were available. Untransformed TN and TP data were used to conduct TITAN analyses. The datasets were subsequently log transformed prior to *post hoc* parametric statistical evaluations so that they met the underlying statistical assumptions of these tests.

Relationships between O/E scores and the TN or TP data were evaluated with simple linear regressions. Next, DWQ's established biological assessment procedures were followed by categorizing each O/E score into one of three categories based on the extent to which models were able to reliably detect departure from reference condition: good (O/E > 0.83; 5% Type I error rate), fair (< 0.83 but > 0.78, 5-10% error rate), and poor (< 0.78, 10% error rate) (DWQ 2010: 35).

The relationships between good and poor sites and TN or TP concentrations (mg/L) were evaluated with logistic regression. Sites categorized as fair (n = 10) were dropped for this analysis because the logistic regression requires binary response data.

Nonparametric deviance reduction (NDR, Qian et al. 2003) models were developed to identify thresholds of TN and TP that best distinguish between good and poor sites (as independently defined by O/E scores). Bootstrapping (10,000 replications, R package boot) was used to calculate 95% confidence intervals (CIs) for each threshold. A two-sample t-test with pooled variance scores was used to determine whether O/E scores above and below TN and TP NDR thresholds were statistically different ($p < 0.05$).

Evaluations of Thresholds Derived from Biological Assessments

The NDR nutrient thresholds were evaluated with receiver operating characteristic (ROC) and relative risk (RR) analyses. ROC analyses confirmed the appropriateness of thresholds in the context of regulatory decisions (Carlisle et al. 2009, McLaughlin 2012,). RR analyses evaluated the extent to which the thresholds identified the extent of risk to stream biota.

Receiver Operating Characteristics

ROC analysis (R package pROC) allowed the identification of thresholds that minimize false positive and false negative assessments as defined by independently derived O/E impairment thresholds (Robin et al. 2011). ROC calculates a single value for model fit called area under the curve (AUC), which is the probability that a randomly chosen response above the threshold will be greater than a randomly chosen response below the threshold. This is approximately the same procedure used in many nonparametric ranked tests such as the Mann-Whitney U or the Wilcoxon rank sum test (Mason and Graham 2002). Error rate estimates (or nonerror rates) are also provided by ROC and are generally more germane to this study than the overall model fit because they relate to the key decision made in the biological assessment process—determining whether a stream is supporting its biological uses.

The analytical basis of ROC is a 2-×-2 error matrix (or confusion matrix) representing two states of condition (impaired or not impaired) and two states of predicted condition (i.e., good versus poor) in relation to a continuous stressor variable (McLaughlin 2012). Error rate statistics were calculated for the proportion of true positives (when nutrient threshold is exceeded and the site is biologically impaired) and the corresponding proportion of true negatives for the range of TN and TP stressors observed among sites. ROC was also used to predict optimal (as defined by the researcher) Type I (false positive) and Type II (false negative) error rates to identify the stressor response thresholds that maximize overall model performance (Hale and Helsthe 2008, Nevers and Whitman 2011).

Relative Risk

RR analyses to identify the threat to aquatic life for sites above and below nutrient thresholds were run with R package spsurvey (v2.2). RR analyses provide an estimate of the relative threat of TN and TP and the 95% CIs surrounding these estimates. RR analysis is commonly used in the medical field and its interpretation is straightforward: What is the factor by which risk increases following exposure to a stressor?

Like ROC, a 2-×-2 error matrix underpins RR analyses. However, in RR, the response and stressor variables must be both categorical and binary. In this case, predefined nutrient-stressor thresholds (i.e., high and low) and the relationship to a binary biological response (i.e., poor or good condition) were examined. The subsequent results indicate the increased risk, relative to other risks evaluated, that the biota will be in poor condition if the stressor exceeds levels of greatest threat to stream biota (Van Sickle and Paulsen 2008). A risk was considered significant if the lower CI was greater than 1.0 for either N or P (Van Sickle and Paulsen 2008, Van Sickle et al. 2006). All analyses were conducted in R v 2.15.0 (R Core Team 2012).

Results

Compositional Changes

Diatoms

Diatom compositional changes could only be evaluated for TP. TITAN revealed thresholds for eight diatom taxa that significantly decreased in abundance and occurrence in response to increasing TP concentrations (Figure 7.1). Taxa were considered significant responders only if purity and reliability were > 0.95. The threshold response for all sensitive taxa—those that decreased with increasing TP—was at a TP of 0.016 mg/L (95% CI = 0.010–0.022 mg/L). On average, the 29 diatom taxa that were statistically tolerant of increasing TP occurred at a TP of 0.042 mg/L TP (95% CI = 0.027–0.047 mg/L). Together, these two thresholds suggest that, on average, the diatom assemblage starts to show appreciable losses of

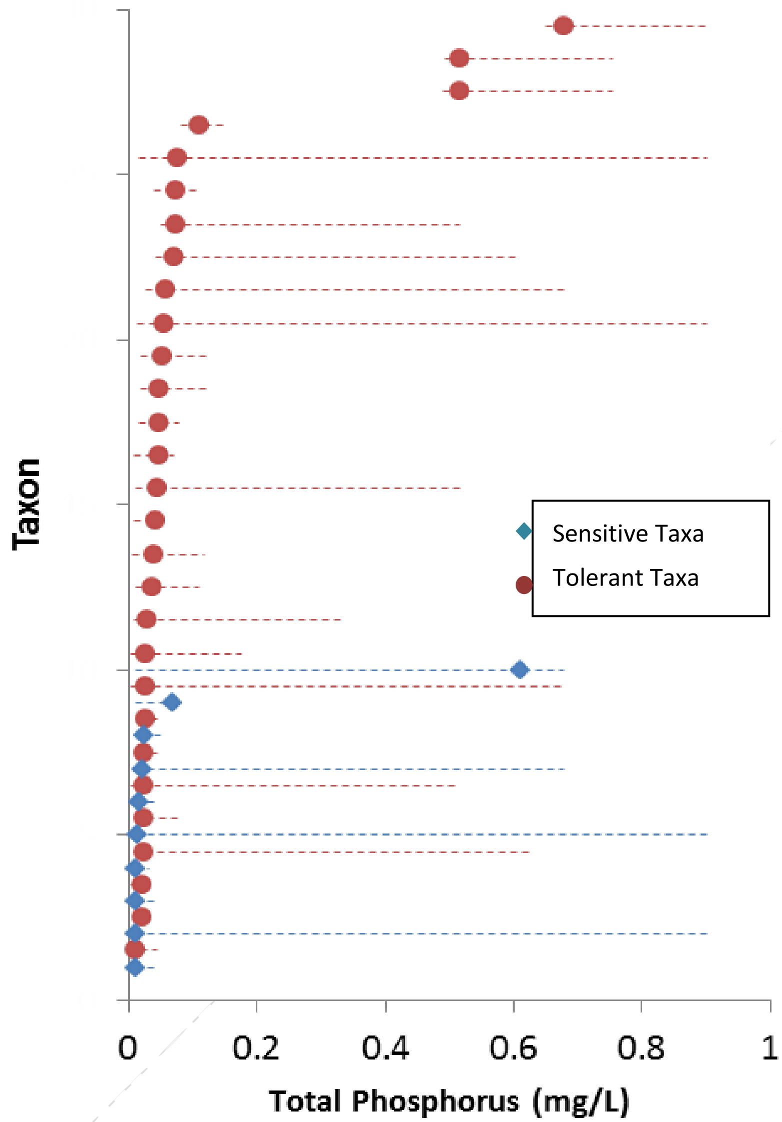


Figure 7.1. Significant indicator diatom taxa plotted in order of their environmental threshold as calculated by threshold indicator taxon analysis. Blue symbols represent sensitive (negative responders) taxa, and red symbols represent tolerant (positive responders) taxa. Dashed lines indicate 5th and 95th percentiles determined by 500 bootstrap replicates.

sensitive taxa at a TP of 0.016, and that more tolerant diatom taxa start to dominate the assemblage at a TP of 0.042 mg/L (Table 7.1). The overall threshold that captures the TP associated with the most appreciable changes in both tolerant and abundant diatom taxa was at 0.022 mg/L (nonparametric change point analysis [nCPA], 95th CI = 0.010–0.047 mg/L). All significant diatom taxon-specific thresholds are presented later in this chapter.

Table 7.1. Thresholds that identify the phosphorus concentrations most strongly associated with appreciable alteration to the composition of stream diatoms.

Diatom Groups	Method	Total Phosphorus (mg/L)		
		Threshold	5^{th} Percentile	95^{th} Percentile
Sensitive	TITAN	0.016	0.010	0.022
Tolerant	TITAN	0.042	0.027	0.051
All	nCPA	0.022	0.010	0.047

Note: nCPA = nonparametric change point analysis, TITAN = threshold indicator taxon analysis.

Macroinvertebrates

With the exception of tolerant taxa thresholds, the TP thresholds obtained from TITAN for macroinvertebrates were similar to those found for diatoms. Significant ($p < 0.05$, calculated from IndVal scores) thresholds were identified for 47 sensitive macroinvertebrate taxa. These taxa significantly decreased in occurrence and abundance as site TP concentrations increased, which resulted in an assemblage-level threshold for sensitive taxa at a TP of 0.011 mg/L (95% CI = 0.003–0.043, Figure 7.2B). Only 24 macroinvertebrate taxa were significantly tolerant of higher TP concentrations; the most appreciable increases in occurrence and abundance occurred at a TP of 0.612 mg/L, which was appreciably higher than the threshold for tolerant diatom taxa. The overall assemblage-level shift from sensitive to tolerant taxa occurred at 0.015 mg/L TP (95% CI = 0.004–0.113mg/L) (Table 7.2).

Overall, TN resulted in more macroinvertebrate taxa with significant increases or decreases than TP. TITAN identified 40 sensitive macroinvertebrate taxa that significantly decreased in abundance and occurrence with increasing TN; 17 tolerant taxa were identified (Figure 7.2A). Additional work on the extent to which temperature, shading, and substrate covary with ecological responses are discussed in Chapter 15 and will be applied during development of site-specific criteria. Significant individual taxa responses were accumulated into an assemblage-level response resulting in a TN threshold for sensitive macroinvertebrates of 0.18 mg/L TN (95% CI 0.14–0.40 mg/L). The TN threshold for tolerant macroinvertebrates was 0.41 mg/L TN (95% CI = 0.36–5.1 mg/L), and the assemblage-level shift from sensitive to tolerant taxa occurred at 0.41 mg/L (95% CI = 0.40–1.1 mg/L TN) (Table 7.2). All significant macroinvertebrate taxon-specific thresholds can be found in Table 7.3.

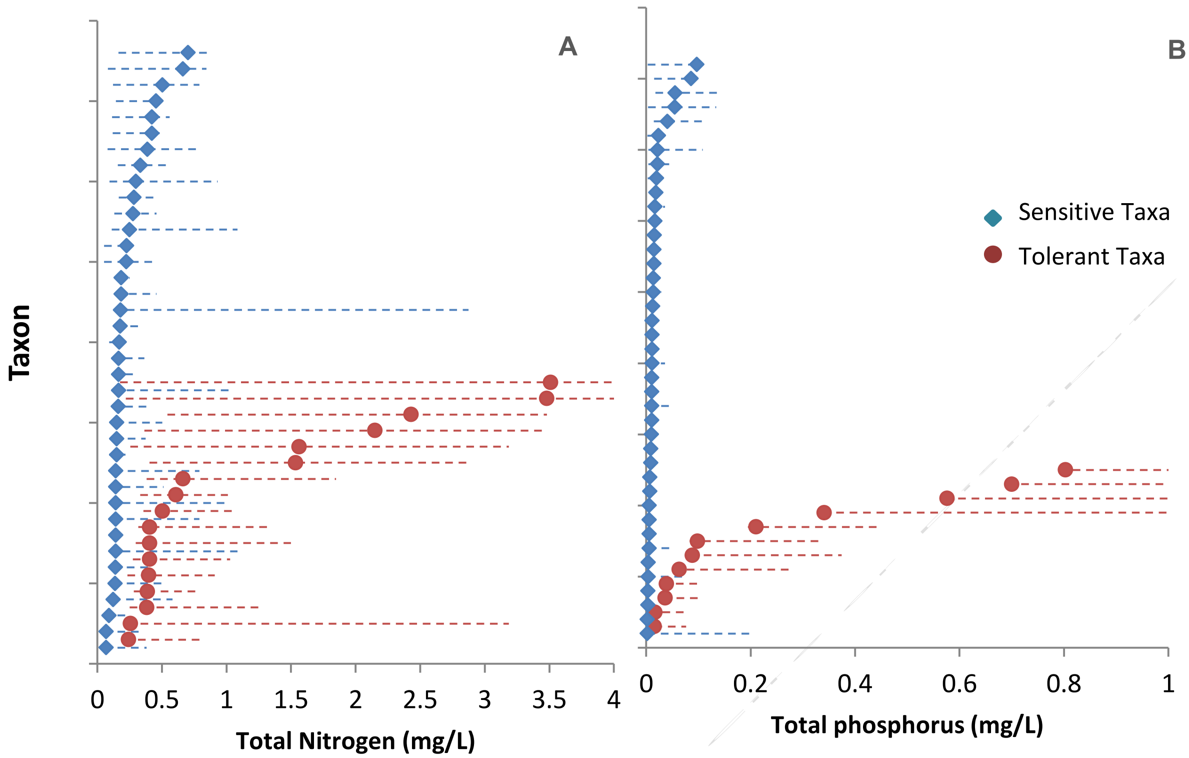


Figure 7.2. Significant indicator macroinvertebrate taxa plotted in order of their environmental thresholds to total nitrogen (A) and total phosphorus (B) as calculated by threshold indicator taxon analysis. Blue symbols represent sensitive (negative responders) taxa, and red symbols represent tolerant (positive responders) taxa. Dashed lines indicate 5th and 95th percentiles determined by 500 bootstrap replicates.

Table 7.2. Thresholds that identify the nitrogen and phosphorus concentrations most strongly associated with appreciable alteration to the composition of stream macroinvertebrates.

Macroinvertebrate Groups	Method	Total Nitrogen Threshold (mg/L)			Total Phosphorus Threshold (mg/L)		
		Best	Percentile		Best	Percentile	
			5th	95th		5th	95th
Sensitive	TITAN	0.18	0.14	0.43	0.011	0.003	0.043
Tolerant	TITAN	0.41	0.36	5.10	0.612	0.042	1.81
All	nCPA	0.41	0.40	0.10	0.015	0.004	0.113

Note: nCPA = nonparametric change point analysis, TITAN = threshold indicator taxon analysis.

Table 7.3. Taxon-specific sensitivity to nutrients derived from threshold indicator taxon analysis (TITAN) models.

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5th Percentile	95th Percentile	Purity	Reliability
Section 1: Diatom Sensitivity to Total Phosphorus										
Anomoeoneis vitrea (Grunow) Ross	0.01	10	1	5.66	0.024	2.56	0.01	0.044	0.96	0.78
Diploneis oblongella (Naegeli) Cleve-Euler	0.01	18	1	10.11	0.008	3.92	0.01	0.901	0.74	0.73
Meridion circulare (Greville) C.A. Agardh	0.01	25	1	13.46	0.004	5.08	0.01	0.043	1	0.95
Cymbella naviculaformis Auersw. Ex Heib.	0.01	10	1	5.96	0.016	2.57	0.01	0.032	0.97	0.85
Synedra delicatissima W. Smith	0.013	10	1	6.92	0.004	4.03	0.01	0.901	0.77	0.74
Hannaea arcus (Ehr.) Patrick	0.015	17	1	7.87	0.02	2.67	0.01	0.045	0.86	0.76
Cymbella affinis Kützing	0.021	152	1	37.38	0.044	2.1	0.01	0.679	0.95	0.87
Achnanthes minutissima Kützing (Achnanthidium)	0.024	247	1	55.82	0.004	8.77	0.022	0.059	1	1
Cymbella microcephala Grunow	0.07	59	1	22.28	0.008	3.24	0.01	0.08	0.98	0.94
Cymbella minuta Hilse ex Rabenhorst (Encyonema)	0.609	157	1	60.41	0.004	3.55	0.01	0.679	0.99	0.98
Amphora perpusilla Grunow	0.01	199	2	52.67	0.004	6.51	0.01	0.045	1	1
Nitzschia inconspicua Grunow	0.021	172	2	59.34	0.004	12.57	0.01	0.03	1	1
Gomphonema clevei Fricke	0.022	39	2	18.02	0.004	5.63	0.018	0.034	0.99	0.98
Achnanthes lanceolata (Breb.) Grunow	0.023	120	2	37.32	0.004	4.8	0.015	0.629	0.98	0.98
Gomphonema parvulum Kützing	0.023	137	2	40.89	0.004	5.46	0.012	0.082	0.98	0.98
Surirella ovalis Brebisson	0.025	99	2	32.2	0.004	4.83	0.021	0.516	0.98	0.98
Nitzschia palea (Kützing) W. Smith	0.025	166	2	47.29	0.004	6.14	0.01	0.049	0.99	0.99
Navicula lanceolata (Agardh) Ehrenberg	0.026	85	2	30.74	0.004	5.63	0.01	0.049	0.99	0.99
Nitzschia tryblionella Hantzsch	0.026	14	2	8.39	0.004	4.1	0.023	0.679	0.97	0.95
Nitzschia hungarica Grunow	0.028	18	2	11.57	0.004	5.51	0.023	0.176	0.99	0.98
Surirella ovata Kützing	0.029	20	2	12.09	0.004	5.23	0.021	0.335	0.99	0.98
Cyclotella meneghiniana Kützing	0.038	64	2	34.93	0.004	10.29	0.026	0.111	1	1

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5^{th} Percentile	95^{th} Percentile	Purity	Reliability
Pinnularia species	0.04	8	2	9.36	0.004	6.89	0.035	0.117	0.99	0.98
Synedra ulna (Nitzsch.) Ehr.	0.041	155	2	51.67	0.004	7.83	0.034	0.047	0.98	0.98
Nitzschia sigmoidea (Nitzsch) W. Smith	0.044	11	2	11.56	0.004	7.18	0.034	0.516	1	0.99
Pinnularia brebissonii (Kutz.) Rabenhors	0.047	6	2	10.34	0.004	8.64	0.04	0.071	1	0.98
Cyclotella species	0.047	12	2	14.24	0.004	7.79	0.03	0.08	0.99	0.97
Amphora coffeaeformis (Agardh) Kützing	0.047	26	2	18.87	0.004	7.02	0.03	0.124	0.96	0.96
Synedra ulna var. *constricta* Venkt.	0.053	11	2	14.03	0.004	8.34	0.035	0.124	1	0.99
Fragilaria brevistriata Grunow (Pseudostaurosira)	0.055	21	2	20.65	0.004	9.05	0.041	0.901	1	1
Bacillaria paradoxa Gmelin	0.057	20	2	20.05	0.004	9.24	0.03	0.679	1	1
Navicula pygmaea Kützing	0.072	9	2	11.35	0.004	5.61	0.03	0.609	0.97	0.96
Navicula capitata Ehrenberg (Hippodonta)	0.075	25	2	20.45	0.004	6.57	0.025	0.516	0.99	0.99
Cymatopleura elliptica (Brebisson) W. Smith	0.075	9	2	14.84	0.004	7.51	0.034	0.111	0.97	0.95
Navicula minuscula Grun.	0.075	28	2	29.57	0.004	9.54	0.059	0.901	1	1
Nitzschia paleacea (Grunow) Grunow in van Heurck	0.111	136	2	49.82	0.008	4.97	0.03	0.15	1	0.99
Nitzschia valdecostata (Lange-Bertalot) Seimonson	0.516	16	2	30.63	0.004	7.36	0.025	0.755	1	1
Navicula tripunctata var. *schizomenoides* (Van Heurck) Patrick	0.516	40	2	40.28	0.008	5.58	0.023	0.755	0.99	0.98
Nitzschia apiculata (Gregory) Grunow	0.679	30	2	51.22	0.004	6.74	0.03	0.901	0.98	0.97
Section 2: Macroinvertebrate Sensitivity to Total Nitrogen										
Empididae	0.068	57	1	76.29	0.004	7.28	0.052	0.382	1	1
Nemouridae, *Amphinemura*	0.068	14	1	39.22	0.004	7.37	0.051	0.323	1	0.98
Dryopidae, *Helichus*	0.091	15	1	39.66	0.004	9.23	0.051	0.219	1	1
Lepidostomatidae, *Lepidostoma*	0.124	50	1	35.73	0.004	4.48	0.119	0.582	0.99	0.98

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5th Percentile	95th Percentile	Purity	Reliability
Brachycentridae, *Micrasema*	0.138	30	1	29.33	0.004	5.86	0.125	0.505	1	1
Uenoidae, *Neothremma*	0.14	10	1	15.1	0.004	5.22	0.124	0.422	1	0.99
Hygrobatidae	0.142	47	1	27.4	0.008	3.43	0.091	1.085	0.99	0.98
Empididae, *Oreogeton*	0.142	5	1	15.15	0.004	9.54	0.091	0.177	1	0.98
Heptageniidae, *Cinygmula*	0.142	26	1	25.18	0.004	5.88	0.125	0.792	1	1
Hydroptilidae	0.142	43	1	31.26	0.004	5.71	0.091	1.011	0.98	0.98
Philopotamidae, *Dolophilodes*	0.142	9	1	13.51	0.004	5.2	0.091	0.514	1	0.98
Rhyacophilidae, *Rhyacophila*	0.142	47	1	30.57	0.004	4.48	0.13	0.792	0.99	0.97
Chloroperlidae, *Suwallia*	0.151	8	1	14.58	0.004	7.48	0.091	0.219	0.99	0.97
Apataniidae, *Apatania*	0.151	12	1	21.14	0.004	8.69	0.132	0.374	1	1
Ephemerellidae, *Serratella*	0.151	14	1	18.57	0.004	5.78	0.138	0.504	1	1
Leptophlebiidae, *Paraleptophlebia*	0.163	38	1	38.58	0.004	9.03	0.111	0.398	1	1
Heptageniidae, *Rhithrogena*	0.165	35	1	24.61	0.004	5.15	0.124	1.05	0.99	0.96
Chloroperlidae, *Sweltsa*	0.165	29	1	36.49	0.004	10.75	0.11	0.276	1	1
Leuctridae	0.166	9	1	15.84	0.004	7.87	0.109	0.365	1	0.99
Dixidae, *Dixa*	0.171	5	1	11.9	0.004	7.33	0.091	0.186	1	0.97
Tipulidae, *Hexatoma*	0.177	35	1	30.46	0.004	6.9	0.151	0.333	1	0.99
Baetidae, *Baetis*	0.179	144	1	53.8	0.012	4.1	0.124	2.874	1	0.99
Ameletidae, *Ameletus*	0.184	25	1	27.2	0.004	7.85	0.14	0.458	1	1
Glossosomatidae, *Glossosoma*	0.184	17	1	23.21	0.004	8.07	0.14	0.249	1	1
Ceratopogonidae	0.223	81	1	38.48	0.004	4.42	0.051	0.424	0.96	0.96
Ceratopogonidae, *Dasyhelea*	0.228	13	1	15.66	0.004	7.08	0.052	0.252	1	1
Heptageniidae	0.25	50	1	32.89	0.004	7.33	0.111	1.085	1	1
Elmidae, *Heterlimnius*	0.276	27	1	27.35	0.004	8.7	0.132	0.458	1	1
Chironomidae, *Micropsectra*	0.283	141	1	54.41	0.004	5.99	0.165	0.454	0.99	0.99
Simuliidae, *Simulium*	0.298	115	1	49.3	0.004	6.68	0.094	0.933	1	1
Ephemerellidae, *Drunella*	0.333	39	1	32.46	0.004	9.16	0.158	0.56	1	1
Perlodidae	0.387	53	1	28.9	0.004	4.44	0.08	0.772	0.99	0.97
Cambaridae	0.422	10	1	10.42	0.004	4.68	0.119	0.483	1	0.98

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5th Percentile	95th Percentile	Purity	Reliability
Chloroperlidae	0.422	37	1	23.63	0.004	4.75	0.115	0.56	0.97	0.97
Heptageniidae, *Epeorus*	0.454	31	1	27.87	0.004	8.55	0.145	0.504	1	1
Perlidae, *Hesperoperla*	0.505	25	1	18.72	0.008	4.32	0.122	0.792	1	0.99
Brachycentridae, *Brachycentrus*	0.664	39	1	26.69	0.004	5.38	0.081	0.846	1	1
Nemouridae, *Zapada*	0.702	39	1	28.4	0.004	6.23	0.163	0.881	1	1
Asellidae, *Caecidotea*	0.242	29	2	24.22	0.004	6.71	0.211	0.818	1	1
Gammaridae, *Gammarus*	0.256	12	2	11.21	0.004	4.18	0.231	3.186	1	0.98
Simuliidae	0.382	40	2	26.6	0.004	5.78	0.249	1.272	0.96	0.96
Leptohyphidae	0.387	11	2	10.89	0.004	4.51	0.283	0.792	1	0.97
Chironomidae, *Eukiefferiella*	0.398	175	2	60.13	0.004	5.1	0.233	0.91	0.98	0.98
Hydrobiidae	0.404	21	2	20.14	0.004	6.93	0.276	1.029	1	1
Hydrobiidae	0.404	38	2	29.74	0.004	7.43	0.298	1.5	1	1
Oligochaeta	0.406	144	2	59.94	0.004	7.86	0.318	1.337	1	1
Erpobdellidae	0.504	26	2	26.55	0.004	9.61	0.357	1.067	1	1
Ephydridae	0.609	10	2	11.41	0.004	5.15	0.333	1.011	1	0.98
Corbiculidae, *Corbicula*	0.664	7	2	9.32	0.008	4.39	0.382	1.848	1	0.97
Physidae, *Physa*	1.536	51	2	54.02	0.004	8.97	0.404	2.874	1	1
Nematoda	1.562	61	2	49.75	0.004	7.08	0.256	3.186	1	1
Hyalellidae, *Hyalella*	2.151	28	2	41.53	0.004	6.32	0.365	3.481	1	0.99
Corixidae	2.43	12	2	34.49	0.004	8.31	0.541	3.481	1	0.99
Planariidae, *Polycelis*	3.481	41	2	48.84	0.008	4.1	0.221	5.079	0.98	0.97
Coenagrionidae	3.511	45	2	58.59	0.004	5.01	0.177	5.079	0.97	0.95
Section 3: Macroinvertebrate Sensitivity to Total Phosphorus										
Hygrobatidae	0.002	47	1	50.51	0.012	3.61	0.002	0.197	0.976	0.91
Leuctridae	0.002	9	1	59.59	0.004	15.11	0.002	0.011	1	1
Empididae, *Oreogeton*	0.003	5	1	35.79	0.004	11.69	0.002	0.01	1	0.986
Ostracoda	0.003	34	1	63.22	0.004	9.83	0.002	0.009	0.974	0.974
Chloroperlidae	0.004	37	1	33.27	0.004	4.49	0.002	0.07	0.964	0.94
Taeniopterygidae	0.004	9	1	36.28	0.004	12.92	0.002	0.008	0.988	0.988

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5th Percentile	95th Percentile	Purity	Reliability
Chironomidae, *Thienemannimyia*	0.006	74	1	49.13	0.004	5.38	0.003	0.053	0.984	0.972
Psychodidae	0.006	26	1	37.52	0.004	8.14	0.002	0.011	0.998	0.998
Philopotamidae, *Dolophilodes*	0.006	9	1	22.83	0.004	9.77	0.002	0.011	1	1
Perlodidae, *Megarcys*	0.006	7	1	26.59	0.004	11.97	0.002	0.015	1	0.998
Apataniidae, *Apatania*	0.007	12	1	50	0.004	17.79	0.002	0.009	1	1
Heptageniidae, *Rhithrogena*	0.007	35	1	43.23	0.004	8.81	0.002	0.017	1	1
Elmidae, *Narpus*	0.009	10	1	16.83	0.004	5.62	0.002	0.028	0.988	0.95
Nemouridae, *Zapada*	0.009	39	1	49.27	0.004	11.03	0.005	0.016	1	1
Ephemerellidae, *Serratella*	0.01	14	1	27.33	0.004	9.51	0.002	0.015	1	1
Heptageniidae, *Cinygmula*	0.011	26	1	41.81	0.004	13.03	0.003	0.016	1	1
Heptageniidae, *Epeorus*	0.011	31	1	40.94	0.004	10.7	0.003	0.051	1	1
Ameletidae, *Ameletus*	0.011	25	1	49.31	0.004	14.78	0.003	0.015	1	1
Perlidae, *Hesperoperla*	0.011	25	1	35.47	0.004	9.46	0.003	0.026	1	1
Elmidae, *Heterlimnius*	0.011	27	1	38.2	0.004	10.44	0.002	0.036	1	1
Capniidae	0.011	23	1	23.02	0.004	5.92	0.002	0.018	0.966	0.96
Leptophlebiidae, *Paraleptophlebia*	0.011	38	1	34.3	0.004	7.05	0.002	0.02	0.992	0.99
Uenoidae, *Neothremma*	0.011	10	1	23.53	0.004	10.31	0.002	0.015	0.998	0.998
Hydropsychidae, *Parapsyche*	0.012	6	1	12.4	0.004	6.38	0.003	0.016	0.994	0.952
Ephemerellidae, *Drunella*	0.014	39	1	44.88	0.004	11.66	0.003	0.029	1	1
Rhyacophilidae, *Rhyacophila*	0.014	47	1	45.59	0.004	11.22	0.004	0.024	1	1
Brachycentridae, *Micrasema*	0.015	30	1	33.06	0.004	8.78	0.009	0.02	1	1
Glossosomatidae, *Glossosoma*	0.015	17	1	27.86	0.004	8.59	0.003	0.017	0.998	0.998
Chloroperlidae, *Sweltsa*	0.015	29	1	43.45	0.004	13.97	0.003	0.021	1	1
Sperchonidae	0.017	80	1	43.36	0.004	5.89	0.004	0.022	0.98	0.978
Elmidae, *Cleptelmis*	0.017	33	1	28.66	0.004	7.1	0.007	0.036	0.978	0.97
Hydropsychidae, *Arctopsyche*	0.019	30	1	31.25	0.004	9.45	0.006	0.031	1	1
Chloroperlidae, *Sweltsa*	0.02	8	1	13.56	0.004	7.31	0.003	0.024	1	0.996
Lepidostomatidae, *Lepidostoma*	0.022	50	1	36.05	0.004	8.45	0.004	0.044	1	1
Pteronarcyidae, *Pteronarcella*	0.022	29	1	17.82	0.012	3.3	0.004	0.108	0.978	0.926

Taxon	Changepoint mg/L	Freq.	Maxgrp	IndVal	Pval	Z score	5th Percentile	95th Percentile	Purity	Reliability
Tipulidae, *Hexatoma*	0.024	35	1	23.58	0.004	5.57	0.002	0.027	1	0.986
Chironomidae, *Diamesa*	0.041	55	1	31.92	0.004	5.62	0.015	0.112	1	0.998
Brachycentridae, *Brachycentrus*	0.055	39	1	21.97	0.004	4.01	0.003	0.134	1	0.99
Baetidae, *Baetis*	0.055	144	1	57.5	0.004	7.68	0.018	0.137	1	1
Elmidae, *Zaitzevia*	0.086	25	1	19.53	0.004	4.45	0.015	0.095	1	1
Ephemerellidae, *Ephemerella*	0.097	32	1	21.32	0.008	3.81	0.003	0.108	0.996	0.986
Hydrobiidae	0.016	38	2	25.92	0.004	4.7	0.012	0.076	0.996	0.996
Haliplidae, *Brychius*	0.017	29	2	21.06	0.004	4.75	0.015	0.074	0.99	0.986
Ephydridae	0.036	10	2	8.49	0.004	3.5	0.02	0.099	0.988	0.914
Erpobdellidae	0.039	26	2	16.41	0.008	3.54	0.009	0.099	0.986	0.948
Leptohyphidae, *Tricorythodes*	0.063	55	2	31.44	0.004	5.02	0.008	0.273	1	0.996
Oligochaeta	0.089	144	2	51.79	0.004	3.8	0.003	0.374	0.98	0.966
Dolichopodidae	0.099	8	2	15.12	0.004	6.89	0.082	0.341	1	0.984
Elmidae, *Microcylloepus*	0.211	37	2	37.25	0.004	6.4	0.036	0.444	0.994	0.986
Planorbidae, *Gyraulus*	0.341	6	2	17.8	0.004	7.04	0.076	1.801	0.968	0.938
Planorbidae	0.576	9	2	35.62	0.004	7.7	0.027	1.801	0.944	0.914
Corixidae	0.7	12	2	43.35	0.004	7.88	0.033	1.801	0.964	0.934
Physidae, *Physa*	0.803	51	2	63.76	0.004	5.72	0.009	1.273	1	0.998
Hyalellidae	1.273	28	2	54.56	0.004	6.29	0.019	1.801	0.998	0.978
Coenagrionidae	1.273	45	2	70.15	0.004	6.2	0.025	1.801	0.99	0.956

Notes: Changepoint is the threshold of total phosphorus (TP, mg/L). Maxgrp 1 indicates sensitive taxa (negative responders) and maxpgrp 2 indicates tolerant taxa (positive responders). Two important diagnostic indicators are calculated from 500 bootstrap replicates. Purity is the proportion of changepoint response direction that corresponds with the observed response direction. Reliability is the proportion of changepoints with an IndVal score that results in significant p-values. Z-scores are calculated by standardizing IndVal scores by subtracting their permutated mean and dividing by their permutated standard deviations. Z-scores may be a better metric than IndVal scores when comparing strength of response between widely distributed taxa and rare taxa.

Biological Impairments

Macroinvertebrate O/E scores decreased with increasing nutrients. A significant, albeit weak, linear relationship was observed among macroinvertebrate O/E scores and TN ($n = 68$, $r^2 = 0.302$, $p < 0.001$) and TP ($n = 243$, $r^2 = 0.294$, $p < 0.001$). The weakness in the relationship may be evidence of other stressors and natural gradients that are expected in structural responses because the effects of nutrients are indirect. For this reason, DWQ has incorporated functional responses with more direct linkages to nutrients into stressor-response (S-R) models directed toward NCC development.

Macroinvertebrate scores were reorganized into binary data (impaired and not impaired) according to DWQ's previously defined biological impairment classes to identify TN and TP concentrations associated with biological impairments. Logistic regression models found both TN and TP concentrations to be significantly related to biological impairment indicator categories (odds ratio = 2.27 [95% CI = 0.81–4.17], z=2.65, $p = 0.008$, and odds ratio=44.51 [95% CI = 31.7–58.99], z=6.42, $p < 0.001$, respectively).

NDR was used as a classification procedure to determine specific thresholds in TN and TP concentrations that best differentiate impaired and not impaired O/E conditions. Sites identified as impaired based on macroinvertebrate O/E scores most frequently occurred at streams with TN > 0.41 mg/L (n = 68, 95% CI = 0.12–0.79 mg/L). This threshold correctly predicted 68% of true positives (prediction probability [PP] =0.68) and 76% of true negatives (PP = 0.76). NDR also identified an average TP threshold of 0.045 mg/L (n = 232, 95% CI = 0.023–0.066 mg/L), which was also quite accurate (true positive PP = 0.65 and true negative PP = 0.89).

Significant differences were found among the O/E scores between the high and low TN sites ($t = 4.22$, $p < 0.001$, Figure 7.3A) and the high and low TP sites ($t = -3.88$, $p < 0.001$, Figure 7.3B) using two-sample t-tests with pooled variances. The significant differences between O/E scores at the high and low TN and TP sites provide further support for the significance of these thresholds.

Alternative Statistical Methods

The strength of the thresholds was evaluated with ROC and RR. ROC revealed that the TN S-R predictions were quite accurate, with a 77.3% chance that a randomly selected site below the TN threshold of 0.42 mg/L will have a higher O/E score than a site above the threshold (AUC = 77.3 (95% CI = 64.2–88.2)). The TP threshold of 0.045 mg/L performed even better with an AUC = 81.4 (95% CI = 75.3–87.3).

All possible TN and TP thresholds were evaluated to identify nutrient concentrations that maximized the percent of both true positives and true negatives. These analyses revealed that the maximum number of both true positives and true negatives occurs at a TN of 0.33 mg/L TN and a TP of 0.045 mg/L TP (Figure 7.4). Depending on the management objective, balancing Type I and II errors may not always be appropriate. Development of site-specific criteria will incorporate additional analyses of potential covariables to reduce both Type I and Type II errors.

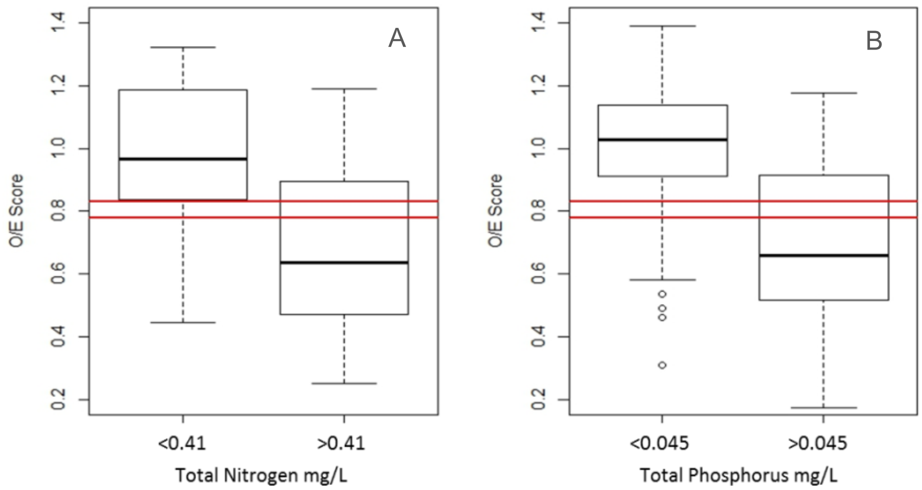


Figure 7.3. Box plots of numeric O/E scores for sites above and below the thresholds determined by nonparametric deviance reduction for total nitrogen (A) and total phosphorus (B). Thresholds were developed from categorical O/E scores between impaired and not impaired sites. The double red lines indicate the range of fair O/E scores. Above the double red lines is considered good (not impaired) and below is considered poor (impaired).

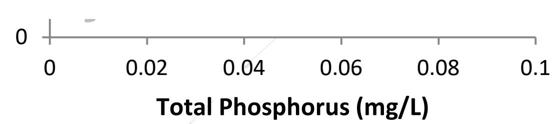


Figure 7.4. Prediction probability curves generated from receiver operator characteristics analysis. Grey line indicates the positive prediction probability (sensitivity) of correctly predicting a biologically impaired site (by O/E score) at a given numeric threshold. The black line indicates the negative prediction probability (specificity). Dashed lines are 95% confidence intervals from 2000 bootstrap replicates.

RR analyses provided another line of evidence that O/E impairments were associated with higher concentrations of both TN and TP. Using RR analysis, the nutrient thresholds established from NDR resulted in an RR for TN of 2.09 (95% CI = 1.57–2.95). This indicates that if the TN threshold is exceeded at a site there is a 2.09-fold greater chance that the O/E score will also indicate impaired conditions compared to a site below the TN threshold. Using the same analysis, an even stronger RR of 3.66 was found for the TP threshold (95% CI = 2.66–5.01).

Discussion

Several measures of macroinvertebrate and diatom assemblages were used to explore associations between changes in the assemblages and varying nutrient concentrations. Results show that specific taxa respond consistently to changes in TN and TP concentrations. Comparing TN and TP to biological assessments demonstrates the clear relationships between increases in nutrients and independently derived measures of biological condition. The risk and ROC analyses helped establish the TN and TP thresholds based on biological assessment metrics in the context of the likelihood of impairments and predicted Type 1 and Type II errors.

Despite the concordance among several lines of evidence, several sources of bias potentially limit the broad applicability of these results. These analyses were conducted using data from different state and federal sources, which potentially biases results by excluding or over-representing environmental gradients. Although the sites are broadly dispersed geographically and all the physicochemical parameters that were evaluated covered the range of conditions observed among randomly selected statewide sites (DWQ, WSA, and NRSA), the potential for site-selection bias remains because DWQ selects sites based on programmatic needs, which often target reference sites (DWQ, WSA, and NRSA) or sites with known or suspected water quality problems. Ideally, a predetermined study design could be developed to better control covariates such as other stressors or natural environmental gradients. This lack of control with sample design may have decreased the strength of threshold-response relationships. Nonetheless, these analyses demonstrate a correlation between high nutrient concentrations and biological degradation. DWQ recognizes that degradation of biological condition relates to stressors and variables other than nutrients. Because this was a statewide exercise, covering large ranges in elevation, watershed size, watershed perturbations, stream size, local habitat conditions, etc., other regional and site-specific covariate stressors, such as temperature, substrate size, stream velocity, shading, etc. could play a significant role in nutrient-threshold determination. Despite these shortcomings, these results, in conjunction with the other lines of evidence presented in this report support the axiom that excess nutrient inputs ultimately degrade stream biota.

Changes in Stream Assemblage Composition in Response to Nutrients

TITAN analysis is built on the estimation of taxon-specific change-points; it evaluates changes in biological structure fundamentally different than more traditional community-based measures of biological condition (index of biological integrity, O/E, etc.). As a result, TITAN models can be built to evaluate biological responses to any environmental stressor. Where multiple stressors co-occur, it is important to evaluate each one to determine the most effective remediation practice. Not all sensitive taxa are equally sensitive to all stressors, so it is also important that TITAN only considers predictable (significant) responders in the calculation of S-R thresholds and for implementation planning to consider all stressors on a site-specific basis. Conversely, differential taxa responses are an intrinsic problem with community-based metrics because taxon-specific responses have the potential to be lost in thresholds based on aggregate, assemblage-level responses.

TITAN generated thresholds for TN and TP, two for each nutrient, that bookend the range of nutrient concentrations that are sufficient to support stream biota. However, DWQ does not aim to protect *all* of the most sensitive species. Even under completely undisturbed natural conditions it is likely that diatom and macroinvertebrate taxa have distributions that vary with physicochemical factors (i.e., floods, droughts, shading, temperature), some of which could either protect against or exacerbate nutrient responses. Allowing nutrient concentrations to achieve levels that result in assemblages dominated by tolerant taxa are likely underprotective because they fail to meet diversity, stability, and biodiversity objectives intrinsic to commonly accepted interpretations of biological integrity, a major CWA objective. TITAN provides the TN and TP concentrations that are, on average, associated with both increases and decreases of individual taxa resulting from nutrient enrichment. A sufficiently protective criterion likely falls somewhere between these two benchmarks, so these brackets provide meaningful context.

Diatoms versus Macroinvertebrates

TP thresholds based on the most sensitive taxa were similar for diatoms and macroinvertebrates (0.016 and 0.011 mg/L, respectively). These concentrations are similar to the best estimates of background concentrations (Chapter 11), which implies that subtle assemblage-level shifts are initiated at fairly low levels of nutrient enrichment. TP thresholds derived from tolerant diatoms remained low (0.022 mg/L), but thresholds for tolerant macroinvertebrates were appreciably higher (0.612 mg/L) than those derived from sensitive taxa. This divergence in assemblage responses is consistent with other investigations that have generally reported diatoms to be more sensitive to nutrient enrichment than macroinvertebrates (i.e., Hering et al. 2006, Justus et al. 2010), although among-assemblage differences can sometimes be obscured by the specific biological metrics or analytical methods used to derive thresholds (Johnson et al. 2006).

Relationship to Biological Impairments

Simple linear regression found a significant, although weak, negative relationship between nutrients and expected macroinvertebrate communities (measured as O/E). When looking at a large range of nutrient concentrations, O/E scores are expected to have a negative relationship with nutrients (like most stressors), yet this this relationship is unlikely to be linear. At very low nutrient concentrations, small increases may actually increase community health by increasing primary and secondary productivity (Hart and Robinson 1990, Mazumder and Edmonson 2002, Slavik et al. 2004). At intermediate levels of nutrient concentration, the negative effects may be masked or exacerbated by site-specific characteristics such as shading, scouring, grazers, etc. (Liess et al. 2009). These sources of variability, left unaccounted for, necessarily obscure linear S-R relationships. Thus, while linear relationships are helpful to evaluate trends, nonlinear and multivariate analyses are almost always required to elucidate nutrient thresholds that are truly detrimental to aquatic communities.

The results of these analyses show a strong relationship between biological impairment and in-stream TN and TP concentrations as determined by logistic regression and NDR. The logistic regression models showed a much stronger fit for TP than TN (odds ratio = 44.5 and 2.3, respectively), although both were significant. The odds ratio can be interpreted as follows: For every one-unit of change in nutrient concentration, the odds of having a corresponding impaired biological condition increases by 44.5% for TP and 2.3% for TN. The difference in magnitude of these odds ratios may be the result of sample size differences (TP, n = 243; TN = 68). Whether TP really has a much stronger relationship to biological impairments is still uncertain, but this analysis demonstrates that an increase in nutrients increases the odds of degradation of the macroinvertebrate assemblage.

NDR enabled further expansion of logistic regression results and provides numeric TN and TP thresholds that best distinguish between sites. On average, O/E scores on either side of TN and TP thresholds differed significantly. NDR identified thresholds that were closer to the higher thresholds determined by TITAN for both TN and TP, which suggests that thresholds based on sensitive taxa responses may be overly protective. These indicators provide another objective metric that can be used to identify streams that are most likely to have nutrient-related impairments for follow-up site-specific confirmation.

Corroboration among Analytical Methods

ROC results confirmed the NDR-derived thresholds by demonstrating that randomly selected O/E scores indicated degraded condition for sites above TN and TP thresholds in 77% and 81% of cases respectively. Perhaps the most interesting ROC insight is the ability to determine Type I and Type II errors at any given threshold (Figure 7.4). If the goal is to maximize model performance then an indicator threshold would be located at the intersection of true positive and true negative prediction probabilities, which equates to 0.33 mg/L TN and 0.045 mg/L TP. If the goal is to minimize Type I errors (false positive impairment conclusions), in order to maximize the chance that resource-intensive site-specific follow-up investigations are focused on real environmental problems, then a threshold that maximized true positive predictions should be emphasized. Alternatively, the importance of false negatives could be emphasized (typically de facto $\alpha = 0.05$, $\beta = 0.20$) if resource managers wanted to err on the side of the Precautionary Principle (e.g., UNEP 1992) and avoid situations where real environmental problems go undetected. In either case, ROC helps inform management decisions by elucidating the tradeoffs involved with pragmatic resource limitations (wasted time and effort addressing false positive, Type II, assessment errors) and fulfilling their mandate to protect water quality (missing impaired waters, Type I, assessment errors).

The relative risk analysis also confirms the results from logistic regression and ROC and demonstrates that increases in nutrients lead to increased probability of impairments. RR is slightly different than ROC—it analyzes the stressor as a binary variable instead of a continuous variable. TN and TP thresholds determined by NDR were used to convert nutrient concentrations into a binary variable of "good" and "poor" based on ambient TN and TP concentrations. The results were similar for logistic regression where TP had a stronger effect than TN, although the relative difference was much smaller. This indicates that increased nutrients indicate an increased probability of biological impairment and that managers can develop a numeric threshold for nutrient concentrations that represent this probability (or risk) of biological degradation without ever knowing the underlying O/E scores.

These additional analyses suggest that nutrient thresholds are not merely a statistical artifact but are based on significant relationships to changes in the composition of stream assemblages. Combining threshold analyses with additional model performance evaluations, such as ROC and RR, has several advantages: improved defensibility of nutrient thresholds, insight into the risk that excess nutrients pose to aquatic life, and the ability to make the most informed management decisions possible. These analyses can be repeated as biological assessment thresholds are developed for additional freshwater assemblages (i.e., fish) to paint a more comprehensive picture of the relative sensitivity of a more diverse range of stream community taxa. Differences and similarities in among-assemblage responses to nutrient enrichment should provide a more complete understanding of impacts to structural aspects of stream condition.

Summary and Recommendations

Macroinvertebrate O/E ratios have been the backbone of DWQ's biological assessment program since its adoption in 2008. Utah's aquatic life uses require the protection of fish (cold water [3A], warm water [3B], and nongame) and other organisms in their food chain. O/E ratios provide quantitative estimates of the extent to which human activities have caused local extinctions of macroinvertebrate taxa, which are a fundamental component of stream food webs. The challenge with using biological assessments is often not in determining if a site is biologically degraded, but the cause(s) of the degradation. The results presented in this chapter suggest that high nutrient concentrations are related to biological impairments (as measured in O/E ratios). Results also demonstrate that thresholds can be developed that identify stream nutrient concentrations that best predict—with known risks and error rates—where detrimental effects to stream biota are most likely to occur. Used individually, these thresholds could not predict nutrient-related impairments, but when coupled with other indicators (e.g., functional responses) biological assessment data can be used to demonstrate that impairments are caused, at least in part, by excessive nutrient inputs.

The regionally derived, structural-based thresholds reported in this chapter describe N or P concentrations that, on average, are associated with alterations in the composition of stream biota. Such regional indicators are useful because they can more accurately identify sites with potential nutrient-related problems. However, management decisions are applied to specific watersheds. Once sites with nutrient-related problems are identified, additional site-specific investigations will need to be designed to more carefully examine cause-effect relationships between nutrients and stream biota. Studies will need to elucidate the relative role of nutrients and other stressors to the loss of stream biota (Chapter 15). These investigations will also need to determine how local habitat conditions (i.e., covariates) diminish or exacerbate the effects of nutrient enrichment on macroinvertebrates or diatoms. DWQ has incorporated tiered monitoring and assessment approaches to accommodate the transition from regional trends to site-specific conditions.

Chapter 8

EFFECTS OF NUTRIENT ENRICHMENT ON RECREATIONAL USES

Key Points

Cultural eutrophication can diminish the recreational value of stream through excessive plant or algal accrual that then creates a physical nuisance to boating and fishing, undesirable odors, undesirable taste in game fish and overall degradation to aesthetics.

A survey was conducted to quantify the extent to which excessive stream algae growth impedes support of recreational uses.

Photographs of streams with benthic algae concentrations that ranged from 50 – 1276 mg chl-*a*/m^2 were mailed to 2700 randomly selected Utah households and participants were asked whether the depicted conditions were desirable or undesirable to their recreation experiences.

The 628 respondents reported a decrease in desirable conditions as benthic algal biomass increased.

Approximately 95%of respondents reported that the depicted stream conditions were desirable until benthic algal biomass reached 150 mg chl-*a*/m^2 where desirability declined to about 60% of respondents. A second decline in desirability ratings occurred at 200 mg chl-*a*/m^2 with about 17% of respondents reporting desirable conditions and then remained consistently low as biomass continued to increase.

Introduction

Most studies that have evaluated the effects of cultural eutrophication have focused on impacts to aquatic life uses, but impacts to recreation uses are also well documented. Increases in nutrients are known to increase rates of microbial production (Cross et al. 2006, Dodds and Cole 2007, Gulis and

Suberkropp 2003), some of which may be harmful to humans, although the epidemiological understanding of specific linkages between nutrients and human pathogens remains poor. Excessive growth of algae and macrophytes can entangle fishing lines, clog boat propellers, and interfere with swimming. Excessive algae growth can also impact aesthetics, degrading the quality of recreational experiences—an issue receiving increasing attention. In Utah and elsewhere, previous negative experiences have been associated with a decline in the number of recreational visits to aquatic ecosystems (CH2MHill 2012, Smith et al. 2015, Suplee et al. 1999).

The impact of nutrient pollution on recreational decisions has been investigated, as has the resulting impact to the economy (CH2MHill 2012, Dodds et al. 2008). A study conducted in 2006 (Hoagland et al. 2002) on the economic impacts of eutrophication determined that harmful algae blooms (HABs) in Maryland's coastal waters resulted in losses of $4 million per year due to recreation and tourism impacts. Another study found direct losses of $10 million to Texas's fishing and tourism industry from a single HAB event (Evans and Jones 2001). Other problems occur when algae blooms affect the taste, odor, or color of water, so recreation impacts from algae blooms are not entirely the result of toxicity. Division of Water Quality (DWQ) recently completed a related study and found that degraded water quality—primarily associated with excess nutrients and water clarity—causes annual losses of $20 million (CH2MHill 2012).

Despite clear ties to recreational use support, DWQ does not have numeric criteria for excessive algae growth; however, Utah's regulatory framework does not ignore algae blooms as a water quality problem. Problems associated with excessive growth have always been part of Utah's water quality standards, which states that waste and other substances cannot be discharged into waters in a way that makes them offensive due to "unnatural deposits, floating debris, oil, scum, or other nuisances such as color, odor or taste," or in a way that results "in concentrations or combinations of substances which produce undesirable human health effects," (Utah Administrative Code R317-2). However, numeric translators that allow DWQ to measure these undesirable characteristics have not been developed, which sometimes impedes the division's ability to address excessive algal growth in streams. Development of numeric criteria for algal biomass requires a more complete understanding of linkages between algae blooms and recreational use impairment decisions—a way to defensibly answer the question: "How much is too much?"

Several recent studies have been successful in quantifying the impact of algae blooms on recreational use decisions using public perception surveys (Smith et al. 2015, Suplee et al. 2009). This chapter describes the efforts of DWQ and collaborators to conduct similar surveys in Utah. It also presents numeric nutrient criteria for protection of recreational uses that DWQ is proposing for headwater streams that were derived from the survey results (Section 2).

Methods

Survey Design

DWQ conducted a public opinion survey to determine whether excessive algae growth alters perceptions of aesthetics and desirable or undesirable recreation conditions. This survey was part of a larger research effort aimed at quantifying the economic benefits of avoiding cultural eutrophication in Utah's waters (see Ch2MHill 2012 for details). The surveys used for these analyses were mailed to 2,700 randomly selected Utah households, and 628 responses were received. The target population for the surveys was every Utah household, so the U.S. Postal Service Delivery System File was used as the sample frame. Survey questions were finalized using the input of three focus groups to help ensure that the survey questions were understood by naïve participants and to eliminate as much bias as possible. To maximize response rates, surveys were sent using a modified Dillman Tillman Tailored Design (Dillman 2011); the design used up to four contact letters, including an advance letter (for additional survey design details see Jakus et al. 2013, Nelson et al. 2015).

Survey participants were provided with two photogrpahs randomly selected from a ppool of eight streams with visible stream bottoms and varying benthic algae cover arranged in a randomly selected order (Figure 8.1). Survey participants were asked to respond to the photographs as follows:

> *Please review the photos of algae in rivers on both sides of the one-page insert included in this survey. For each photograph on the insert tell us if the level of algae would be desirable or undesirable for YOUR most common uses of rivers, if any. There are no correct answers; this is your opinion only.*

For each stream in the photographs, the Montana Department of Environmental Quality (DEQ) had previously quantified the extent of algae growth with per-area chlorophyll-*a* (chl-*a*) concentrations that were obtained from a composite of 10–20 replicate algae samples (Suplee et al. 2009). The benthic chl-*a* concentrations in these streams ranged from < 50 mg chl-*a*/m^2 to 1,276 mg chl-*a*/m^2, in intervals of ~50 mg/m^2 (Suplee et al. 2009). Importantly, survey participants were not sent any data associated with the photographs, so variation among responses was entirely related to visual differences among the photographs.

Analysis and Interpretation of Survey Results

The consistency and representativeness of survey results was evaluated in several ways. Nonresponse bias was evaluated using propensity score adjustment and 12 different demographic parameters (see Jakus et al. 2013: Appendix D) to determine whether households that responded differed in any way from those who did not (Groves 2006). Ordered probit models were then used to determine whether differences among those who did respond could be attributed to demographic or stated preferences and experiences. Finally, monotonicity tests were conducted to evaluate the extent to which each respondent was consistent in their evaluations. In other words, once they stated that conditions in a

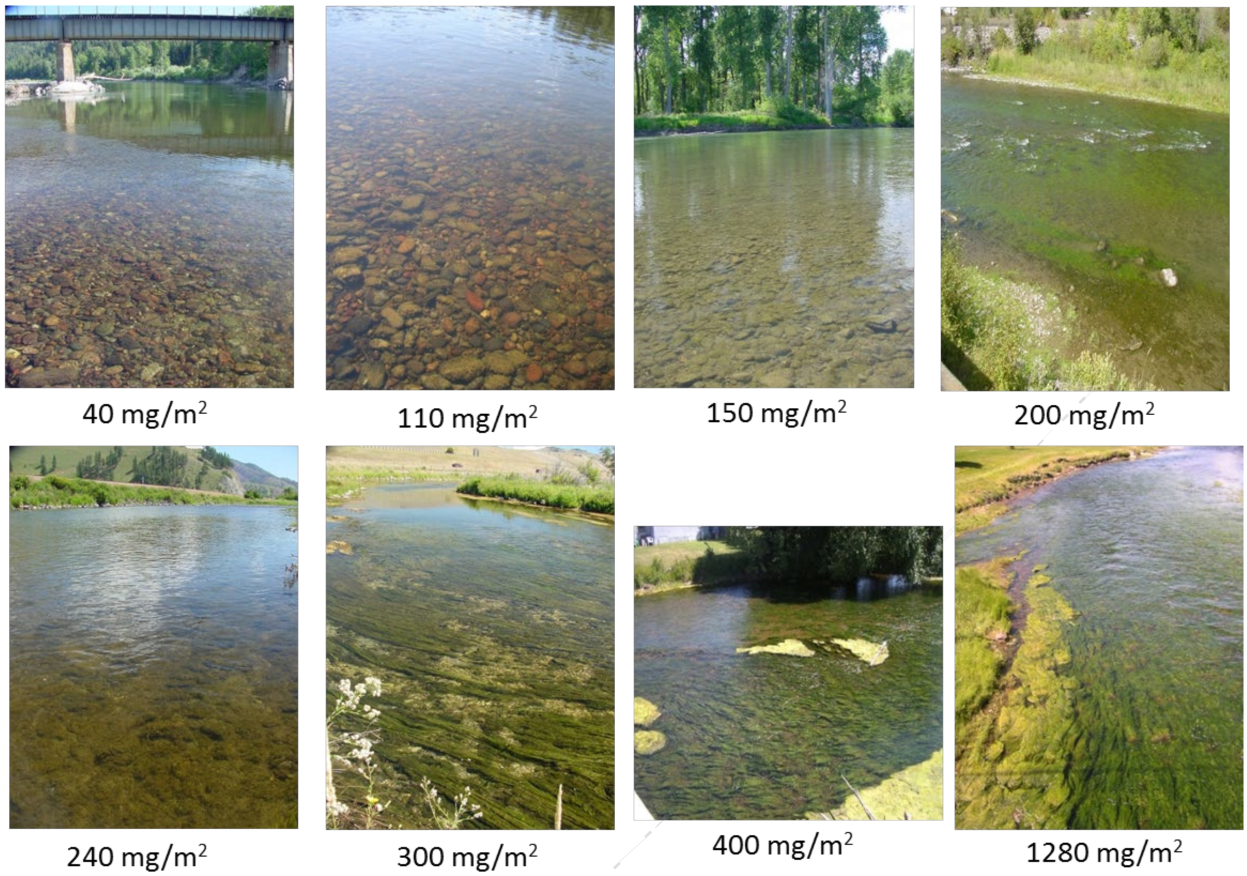


Figure 8.1. Photographs provided to survey participants and their associated chlorophyll-*a* concentration (which were not provided to survey participants); photographs and data were obtained from Suplee et al. 2009.

photograph were undesirable, did they respond similarly to all photographs that depicted higher levels of algal biomass?

Once received, survey responses were associated with the blind quantitative measures of algae cover. The relationships among benthic chl-*a* concentrations and percent desirable conditions were analyzed with a Spearman rank correlation. Differences between user responses (those who recreate at streams) and nonuser responses were evaluated with a Kruskal-Wallis rank-sum test. Differences and similarities between the Utah and Montana survey results were explored. All analyses were conducted in R (R Core Team 2012).

Results

A total of 625 completed surveys were received, which equated to a response rate of 25.3%. However, some responses did not answer all the questions, so the results presented here were derived from the 555 responses where preferences were provided for all 8 photographs. Propensity scores were insignificant for all the demographic data evaluated, indicating that those households who responded were not significantly different than those who did not in attributes like education, income, or geographic location. About 80% of households provided responses that were weakly or strongly monotonic, meaning that once they reported that an algal biomass was undesirable, they generally continued to answer similarly at increasingly higher concentrations of algal biomass (Jakus et al. 2017).

The survey revealed a strong and consistent relationship between benthic chl-*a* concentrations and what the public viewed as desirable conditions (Figure 8.2A). The percent of respondents who indicated desirable conditions decreased from 96.6% at streams with benthic chl-*a* concentrations of 40 mg/m^2 to a low of 7.8% at streams with 1280 mg/m^2 (Table 8.1).

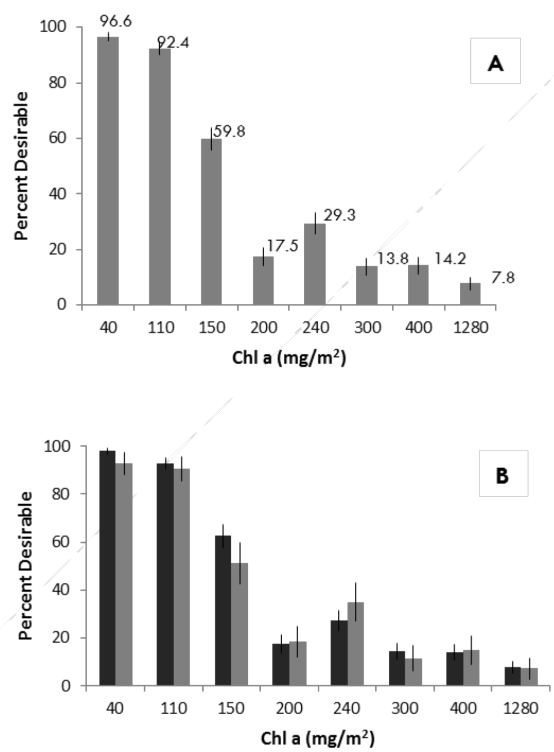


Figure 8.2. Percent of all survey respondents who indicated that streams of varying algae concentrations were reflective of desirable conditions (Panel A). Also depicted is the distinction between respondents who recreate at streams (Panel B, black bars) and those who do not (Panel B, gray bars).

Table 8.1. Percent desirable survey responses between two user groups (users and nonusers) from Utah's survey and the Montana survey.

Photograph	Chlorophyll-*a* mg/m^2	% Desirable, Utah		% Desirable, Montana	
		Nonusers	Users	Nonusers	Users
1	40	92.8	98.0	95.6	98.2
7	110	90.6	92.9	94.9	93.6
6	150	51.2	62.5	69.7	75.8
5	200	18.3	17.4	16.5	31.8
2	240	34.9	27.1	28.8	29.1
8	300	11.4	14.4	12.6	20.2
3	400	14.7	14.1	16.7	11.5
4	1280	7.1	7.8	11.3	9.1

Source: Suplee et al. 2009
Note: Responses between user groups and between the two states did not differ (analysis of variance p=0.94).

A significant negative correlation among benthic chl-*a* concentrations and percent desirable responses was identified (Spearman's r=-0.95, $p < 0.001$). Survey responses demarcated two distinct thresholds of benthic algae concentrations. As benthic algae increased from 110 to 150 mg/m^2 the percent of "desirable" responses declined from 91 to 61%. The second threshold occurred at the next incremental increase in algae cover (from 150 to 200 mg/m^2) where desirable responses fell from 61% to 18%. Desirable responses remained consistently low for all subsequent incremental increases.

No significant differences were found between water-based recreationists (users) and nonusers (Figure 8.2B). The results were nearly identical to those obtained from a similar survey conducted in Montana (Kruskal-Wallis rank-sum test, p = 0.94, Table 8.1; Suplee et al. 2009). Examination of individual photographs also revealed remarkable similarities between the two states. A statistically significant difference (p = 0.002) between the two states was observed for only one photograph (150 mg chl-*a*/m^2); Montana respondents indicated a slightly stronger tendency than Utah respondents to say this stream was desirable.

Discussion

This survey revealed remarkable consistency in the algae conditions that Utahns consider undesirable. The original hypothesis was that users would be less tolerant that nonusers due to their increased familiarity with similar waters; however, this was not the case. The similarity in responses from both groups suggests that these results are a broad reflection of the opinions of Utah citizens. The representativeness of these results is also bolstered by the marked similarity between the Utah and Montana surveys (Table 8.1) and suggests that there may be broad consensus about desirable stream conditions in the intermountain west.

The two thresholds provide some insight into benthic algae densities that may be protective of recreational uses. The higher threshold captures a drop to baseline conditions and clearly represents degraded aesthetics. In contrast, the first threshold captures the initiation of decline in aesthetics and is therefore closer to values that are potentially protective of recreation uses. These thresholds are supported by other studies. For instance, based on a review of several investigations, Dodds and Welch (2000) concluded that undesirable algae densities generally fall between 100 and 200 mg chl-*a*/m^2. Biggs (2000) stated that once chl-*a* biomass is > 150–200 mg/m^2 it could be reasonably concluded that the conditions were "probably unnatural" and that they likely compromised contact and sports fishery recreational uses. Such consistency among all these studies suggests that these thresholds are not merely an artifact of DWQ's study design or the specific images shown to survey participants. Benthic algae blooms of this magnitude almost always require both excessive nutrients and habitat that is favorable to algae growth (i.e., sufficient light, sufficient length of time between scouring floods).

Utah has traditionally protected recreation uses exclusively using *E. coli* cell counts and only recently proposed to assess lakes and reservoirs using harmful algae bloom data. Both methods are directly tied to human health threats. These survey responses provide the information necessary to expand protection of recreation to include the protection of aesthetics.

The ability to quantitatively relate nutrient enrichment to a loss of aesthetics is important because degradation of aesthetics is potentially more likely to alter future recreation decisions than those related to pathogens. People are generally unfamiliar with pathogen concentrations at specific waters, even those where problems have been documented. In contrast, conditions that can be seen by everyone and that are consistently viewed as "undesirable" are more likely to alter future decisions about where to recreate. Assessments of recreational use support based on both health-related and aesthetics more completely cover two different, but equally important, stressors that can degrade recreational uses.

Proposed Numeric Criteria for Recreational Uses

Water quality standards require that the uses be protected, which means that water quality standards should set levels that preclude impairment from occurring. In this study, people started to identify streams as representing undesirable conditions as chl-*a* levels increased from 110 to 150 mg chl-*a*/m^2, with a major decline at 200 mg chl-*a*/m^2. DWQ proposes a criterion of 125 mg chl-*a*/m^2 to protect against declines in recreation use; this standard is the equivalent of 49 g-C (ash free dry mass).

SECTION 2

A Framework for the Development of Numeric Nutrient Criteria for Utah's Headwaters

Chapter 9

PRIORITIZATION OF HEADWATER STREAMS FOR NUMERIC NUTRIENT CRITERIA DEVELOPMENT

Several years ago, the Division of Water Quality (DWQ), in collaboration with stakeholders, made a decision to focus initial efforts on regional numeric nutrient criteria (NNC) development on headwater streams. This section of the technical support document (TSD) provides the results of two headwater-specific investigations to help DWQ achieve this objective. First, several classification analyses were conducted using ambient nutrient data collected from headwater streams and a suite of GIS-derived environmental gradients to determine whether additional categorization of headwater streams had the potential to minimize natural variation of stressor-response (S-R) relationships (Chapter 10). Second, an analysis was conducted with all the nutrient data collected from Utah's headwater streams over a nine-year period (Chapter 11). This exploration was conducted using nutrient data obtained from all headwater streams and also from a subset of reference sites. The first set of analyses allowed DWQ to estimate the extent of enrichment in Utah's waters, whereas the second analysis allowed DWQ to use frequency distribution methods to obtain an additional benchmark for the derivation of headwater NNC. The results of these investigations are combined with information gleaned from the S-R investigations (Chapters 2–9) to propose and then test headwater NNC, which are described in Section 3 of this TSD.

Key Points

Headwater streams are critically important sources of water, and the protection of these ecosystems is needed to maintain the quality of life of Utah citizens and ongoing vitality of many sectors of Utah's economy.

DWQ and stakeholders representing diverse water quality interests prioritized headwater streams for the development of numeric nutrient criteria (NNC).

This section of the report decribes the technical details needed to apply the previously described stressor-response relationships to develop headwater NNC.

The principal reason for the prioritization of headwater streams for NNC development is that they are critically important ecosystems—both ecologically and economically. Ecologically, these streams contribute to the biological integrity of all streams by providing critical hydrological connectivity among streams across large landscapes (Freeman et al. 2007). At regional scales headwater streams are critically important for the maintenance of aquatic biodiversity (Clarke at al. 2008), in part because they are physically diverse with a corresponding diverse breadth of niches (Lowe and Likens 2005). Native fish, like Utah's cutththroat trout (*Oncorhynchus clarki*), inhabit these streams year round or migrate to these streams in early spring for spawning. In an economic context, headwater streams provide many important ecosystems services. Headwaters, which represent approximately two-thirds of total river miles in Utah (Freeman et al. 2007), protect downstream waters through nutrient retention (Bernhardt et al. 2005), maintenance of sediment transport (Lowe and Likens 2005), and organic matter storage and processing (Muotka and Laasonen 2002). In Utah, the majority of water falls as mountain snow, so headwater catchments are a critical part of water storage. For over three decades DWQ has acknowledged the importance of headwater streams and afforded them Antidegradation Category 1 or 2 protections (Figure 9.1; Utah Administrative Code [UAC] R317-2), which precludes discharges (Category 1) or requires that discharged pollutants do not exceed background concentrations.

Despite these protections and similar protections implemented elsewhere, headwater streams remain threatened (Myer et al. 2007). Two important and interrelated threats to these ecosystems are habitat degradation and anthropogenic nutrient enrichment. Finlay (2011) reviewed metabolism data collected from over 200 streams and found that primary production in human-influenced headwater streams was higher than comparable reference sites, with most human-influenced streams 600% higher than reference sites. Habitat degradation can exacerbate nutrient effects. Intact riparian conditions buffer the effects of nutrient enrichment directly by decreasing photosynthetic active radiation (PAR) and indirectly by maintaining habitat complexity (Greenwood and Rosemond 2005). Because unaltered headwater streams are typically nutrient poor, resident biota are adapted to these conditions and are often relatively sensitive to nutrient enrichment (Miltner and Rankin 1998). Overall, incremental degradation of headwaters is more likely to have deleterious effects on these ecosystems relative to larger streams.

For purposes of NNC development, DWQ has elected to define the population of applicable headwater streams as perennial streams encompassed within Antidegradation Category 1 and 2 waters (UAC R317-2-12; Figure 9.1). These streams consist of waters that Utah's Water Quality Board has determined are "…of exceptional recreational or ecological significance or have been determined to be a State or National resource requiring protection." Additional protections for these waterbodies include prohibiting all point source discharges to Category 1 streams and limiting pollutants discharged to Category 2 waters to background conditions. Most of the stream segments with these protections are within U.S. Forest Service boundaries, which encompass approximately 8.2 million acres, over 15% of the acreage in Utah (Gorte et al. 2012). All told, about 8,000 miles of perennial streams are within these watersheds, which is about 47% of the total perennial stream miles in Utah.

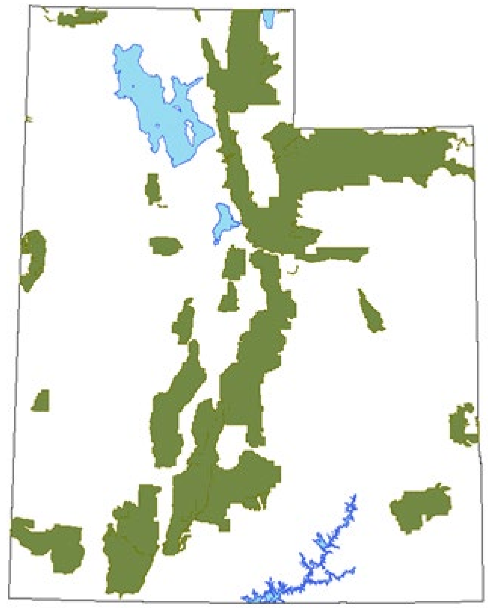


Figure 9.1. Utah's Antidegradation Category 1 and 2 boundaries are shown in green. In these waters the state does not allow point source discharges (Category 1) or allows only discharges equal to background concentrations (Category 2).

Section Organization

This section of the TSD contains a proposal for criteria for headwater NNC and materials supporting the selection of those criteria. It begins with a summary of ambient nutrient concentrations observed among headwater streams, followed by a detailed explanation of proposed NNC elements and their underlying rationale. The section also presents the results of an investigation conducted to test the veracity of proposed NNC elements.

Chapter 10 provides the results of a classification investigation conducted to determine whether additional subclasses of headwater streams were needed to better address natural variability in ambient nutrient concentrations. Chapter 11 presents data on the distribution of ambient nutrient concentrations among headwater streams so that the proposed criteria can be examined against the NNC criteria that would be generated using frequency distribution methods. This includes a summary of nutrient concentrations observed among headwater reference sites to use as an additional benchmark to inform appropriate nutrient concentrations for the proposed NNC. The range of nutrients among all headwater streams is also explored for reference sites and sites where human-caused enrichment has occurred. This is done to better understand the extent of enrichment that has already occurred and to estimate the extent of impairments that might result from NNC adoption. The proposed NNC and the results of the conformational investigation are presented in Section 3.

Chapter 10

EVALUATION OF POTENTIAL SUBLASSIFICATIONS OF HEADWATER STREAMS

Key Points

Comparing streams with similar attributes minimizes natural variation in ambient nutrient concentrations and improves the accuracy of stressor-response relationships.

A study was conducted to determine if further classification of headwater streams had the potential to improve headwater numeric criteria.

Environmental attributes know to be associated with natural variation in ambient nutrient concentrations were obtained for all headwater watersheds.

A cluster analysis (k-means, k = 2) was conducted to divide the streams into two groups that were as physically distinct as possible.

Neither nitrogen nor phosphorus differed between the two groups of streams, suggesting that headwater streams do not require further classification.

Introduction

Addressing natural variation for both background nutrient concentrations and ecological responses is a central challenge of numeric nutrient criteria (NNC) development. Background nutrient concentrations vary as a result of naturally occurring physical and environmental conditions such as the mineral composition of soils and bedrock, soil erosion rates, organic matter inputs from watershed runoff, and physical determinants of nutrient resident time such as discharge and channel gradient (Smith et al. 2003). The variation resulting from these characteristics can be considerable. Ambient stream nutrient concentrations among reference sites in different regions of the United States vary by two orders of magnitude (Lewis et al. 1999, Clark et al. 2000). Environmental gradients can buffer or exacerbate ecological responses to nutrient enrichment, which means that the specific nutrient concentration that is protective of aquatic life varies among streams (Dodds and Welch 2000).

Classification minimizes natural variation by systematically grouping streams with similar physical and environmental characteristics. If classification is successful, the variation in ambient nutrient concentrations among streams within each class of streams is minimized. Ideally, natural variation in within-group ecological response is similarly minimized because streams within each group may also have similar ecosystem structure and function. Limiting among-stream comparisons to ecologically similar streams can also improve the strength of stressor-response (S-R) relationships, which results in NNC that are more accurate and appropriately protective of designated aquatic life uses (USEPA 2000). As a result, classification is typically among the first steps conducted to derive NNC.

In 2008, ambient nutrient data collected from rivers and streams throughout Utah were analyzed to determine whether there was sufficient data to perform an S-R analysis between nutrients and alterations to the composition of macroinvertebrate assemblages of wadeable streams (Paul 2009). As part of this work, several different classification schemes were evaluated. One was an *a priori* classification scheme that compared differences in ambient nutrient concentrations and several macroinvertebrate responses among streams within Omernik level III ecoregions. Other approaches included development of empirical models that used data collected at reference sites to predict ambient nutrient classes. These analyses did not produce a classification scheme that minimized among-group variance in nutrients or responses on a statewide basis. However, these analyses did identify several limitations in the underlying data, which Division of Water Quality (DWQ) has subsequently tried to address. In addition, although these preliminary analyses were conducted on all types of streams in Utah, the initial focus for NNC development, which is presented in this technical support document, is on headwater streams.

This chapter presents the results of several analyses conducted to determine whether headwater streams required further classification into smaller groups for purposes of NNC development. To make this determination, a series of watershed-scale classifications were made using GIS-based measurements of natural environmental gradients (e.g., channel gradient, lithology, climate) to create subclasses of headwater streams that were as similar as possible in multivariate space within groups while also maximizing among-group differences. Between-group differences in ambient nutrient concentrations were then examined to determine whether subclasses of headwater streams were needed for purposes of NNC development.

Methods

Classification of All Headwater Streams

Watershed Delineation

Relatively small watersheds were used as the experimental units because differentiating headwaters from streams lower in watersheds is already an *a priori* classification, and the focus of this investigation was whether finer-scale classifications were warranted. To identify suitable watersheds for this analysis, all 12-digit hydrologic unit code (HUC) watersheds that are currently protected as

Antidegradation Category 1 and 2 streams were selected. HUC-12 watersheds that overlapped Antidegradation Category 1 and 2 waters were also included (Figure 9.1).

Environmental Gradient Descriptors

Data for each watershed describing numerous physical and environmental characteristics—environmental gradients—that others have found useful in accounting for landscape-level differences in ambient nutrient concentrations were compiled for all headwater streams where ambient nutrient concentration data had been previously collected. The analysis primarily focused on environmental gradients that could be quantified using readily available GIS data (Table 10.1). Geographic gradients of slope and elevation were obtained from digital elevation models (DEMs). Soil information from the Natural Resource Conservation Service's (NRCS) state soil geographic database (STATSGO) and lithology gradients that quantify background nitrogen (N) and phosphorus (P) within soils and bedrock were also included (Olson and Hawkins 2013). Climate data from the physiographically sensitive mapping of temperature and precipitation (PRISM) database maintained by Oregon State University was used in the analysis because background precipitation alters vegetation composition and stream hydrology. The scale of measurement for these environmental descriptors varied considerably, which can artificially under- or over-weight them in classification analyses, so each variable was normalized by subtracting the population median from each observation and then dividing by the population standard deviation prior to subsequent analysis.

Table 10.1. The relative importance (component loading) of environmental attributes used to classify streams into two physically distinct groups generated from both broad and refined principal components analysis models.

Parameter Code	Data Sources	Variable Description	Component Loadings			
			All Variables Model		Refined Model	
			Factor 1	Factor 2	Factor 1	Factor 2
TMEAN	PRISM	Predicted annual mean monthly air temperature (^{0}C)	0.910	0.089	0.958	-0.180
LST32	PRISM	Mean day of year for last freeze	-0.875	-0.032		
FST32	PRISM	Mean day of year for first freeze	0.854	-0.043		
elev		Mean watershed elevation (m)	-0.788	-0.215	-0.846	-0.261
Mean P	PRISM	Predicted annual mean precipitation (mm)	-0.781	0.121	-0.726	0.171
XWD	PRISM	Predicted annual number of days with precipitation	-0.769	0.324		
weg	STATSGO	Wind erodibility group	-0.606	-0.176		
rockdepth	STATSGO	Depth of soil to bedrock (inches)	-0.564	0.170	-0.547	0.274
awch	STATSGO	Available water capacity of soils (fraction)	0.209	0.839	0.370	-0.875
bdh	STATSGO	Soil bulk density (g/m^{2})	0.124	-0.834		

Parameter Code	Data Sources	Variable Description	Component Loadings			
			All Variables Model		Refined Model	
			Factor 1	Factor 2	Factor 1	Factor 2
perm	STATSGO	Precipitation permeability of soils (inches/hr)	0.112	-0.790	-0.083	-0.806
kfact	STATSGO	Soil erodibility factor	0.492	0.611	0.560	0.621
om	STATSGO	Organic matter content of soils (%)	-0.483	0.353		
WTAVGP	USU	Total phosphorus by weight in bedrock (%)	0.141	0.165		
WTAVGN	USU	Total nitrogen by weight in bedrock (%)	0.266	0.126		
slope	DEM	Watershed-scale channel gradient (%)	-0.280	0.116		
WTAVGPERM	USU	Water permeability of bedrock (µm/second)	0.178	-0.087		
tfact	STATSGO	Mean soil loss tolerance factor	-0.397	0.070		
HVDR	USU	Hydrologic variation (mean monthly Q/mean maximum Q)	0.073	0.055		
wtdepth	STATSGO	Water depth in soils (feet)	0.125	-0.054		

Notes: Watershed-scale attributes were combined from several sources, including: NRCS state soils geographic database (STATSGO), USGS DEM, and lithology maps compiled by the Western Center for the Monitoring and Assessment of Freshwater Ecosystems (USU).

Classification Analysis

The classification analyses were iterative; they started with a k-means cluster analysis (k = 2) that used all possible environmental descriptors to identify two groups of streams that were as distinct as possible. Next, two–factor principal components analysis (PCA) was used to determine the relative importance of each of the 20 environmental gradient variables based on their component loading scores (contribution to each component). To control for colinearity and identify the most parsimonious solution possible, highly correlated ($r^2 > 0.5$) environmental attributes were identified, and then the environmental characteristic with the highest PCA component loading was selected from the correlated attributes. Following this variable reduction exercise, a second k-means cluster analysis was conducted to identify two groups of streams that best minimized within-group variability and maximized between-group variation in environmental space.

Nutrient Data

The primary analytical focus was on whether N and P concentrations differed between these two groups of environmentally distinct streams. However, before these analyses could be conducted, some data artifacts created by historic laboratory methods had to be addressed. In the past, DWQ analyzed samples for dissolved inorganic N (DIN, nitrate + nitrite + ammonia) instead of total N (TN). For these classifications, DIN was used instead of TN, so that the classifications could be based on the largest number of sites possible. Also, until recently DWQ's nutrient lab analysis had detection limits that were relatively high, which is problematic at reference sites where nutrient concentrations are low. To address the resulting nondetects, the nonparametric Kaplan-Meier (K-M) method of survival analysis was used (R package NADA; Lee 2013, R Core Team 2012). The K-M method was developed to extrapolate right-censored survivor data, so the K-M "backward" method was used to extrapolate nondetects from left-censored laboratory data to obtain values between laboratory reporting limits and zero.

Evaluation of Potential Sub-classification Groups

Once subclassifications were defined using environmental gradients, between-group differences in DIN and total P (TP) concentrations were evaluated. Of the 99 candidate reference sites, 6 were dropped because they were not located in watersheds that intersected with Antidegradation Category 1 and 2 boundaries. Another 8 sites fell within the same 12-digit HUC, so the nutrient concentrations of these sites were averaged. This initial screening effort resulted in a total of 89 unique 12-digit HUCs where ambient nutrient concentrations were available for one or more streams. Next, each HUC was coded to its k-cluster group and then between-group differences in the average DIN and TP observed at each location were evaluated. A two-tailed Peto-Prentice test (NADA package) was used to determine whether DIN or TP was statistically distinct between the classes because this method is relatively unbiased by larger numbers of nondetects and the subsequent resampling that were required for these data.

Results

Classification

Relative Importance of Environmental Gradients

From the initial two groups established with k-means cluster analysis and all 20 candidate environmental descriptors, several watershed characteristics were identified that were particularly important in describing between-group differences for headwater streams (Table 10.1, Figure 10.1). These key attributes included: annual mean predicted air temperature (TMEAN), elevation (elev), annual mean predicted precipitation (MEANP), annual mean of predicted number of days with precipitation (XWD), depth of soil to bedrock (rock depth), available water capacity of soils (awch), permeability of soils (perm), and a soil erodibility factor (kfact) (Figure 10.2). PCA axis loading revealed that the first axis, which captures the largest portion of among-site variation, primarily included characteristics associated with elevation and weather (temperature and precipitation). The second PCA axis described watershed attributes associated with soil characteristics (permeability and erosion).

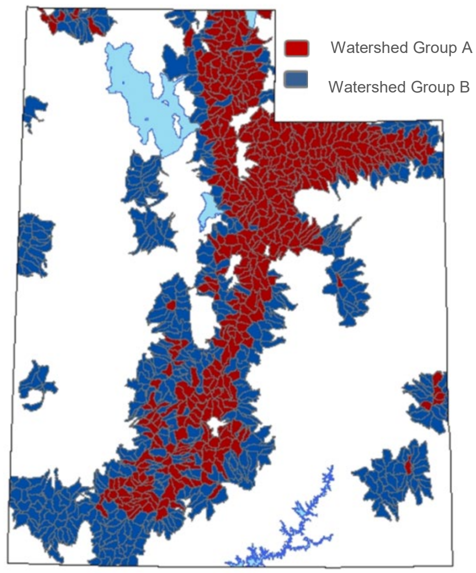


Figure 10.1. All 12-digit hydrological unit codes used in ecological classification grouped by results of k-clustering using best-performing variables.

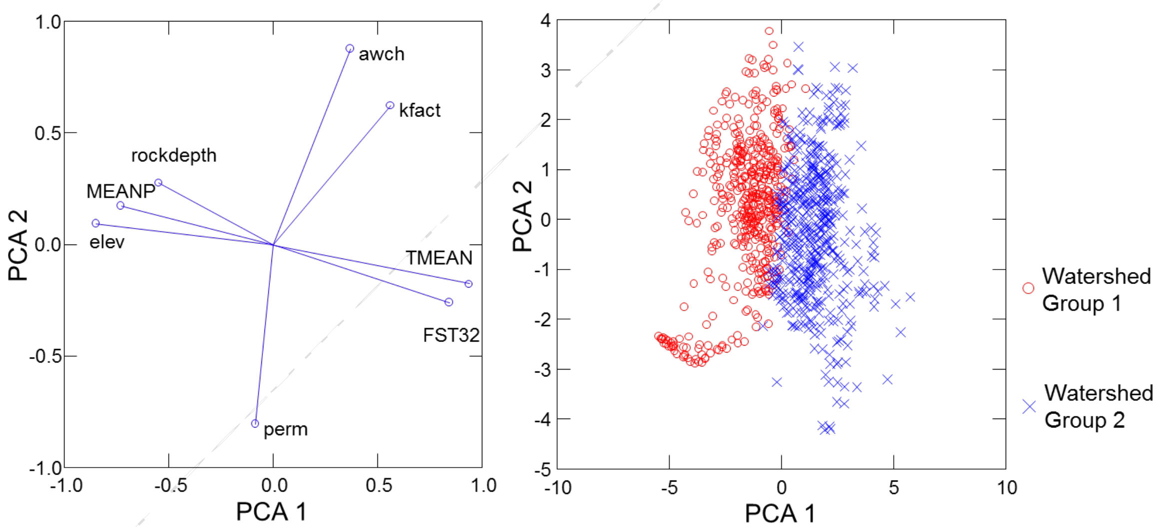


Figure 10.2. Results of principal components analysis for determining physical and environmental factors that best classify 12-digit hydrological unit codes into two distinct groups. This analysis was run with only the best-performing variables from the entire population (see Table 10.1) determined by component loadings.

Among-Group Differences in Ambient Nutrient Concentrations

Nutrient concentrations at reference sites contained a relatively large number of nondetects. About 41% of historic DIN results for these sites were below the reporting limit of 0.075 mg/L. TP chemical analyses were equally problematic with ~61% of reference samples falling below the reporting limit of 0.02

mg/L TP. All DIN and TP nondetects were subsequently censored by distributing values between reporting limits and zero (R, NADA package) for purposes of these analyses.

As expected, the reference sites had fairly low nutrient concentrations. The censored data resulted in a mean reference DIN concentration of 0.192 mg/L (95% confidence interval [CI] = 0.125–0.259) (Table 10.2). TP was also low at these sites, with a population mean concentration of 0.017 mg/L (95% CI = 0.013–0.022). There were no significant differences between watershed groups for both DIN (p = 0.906) and TP (p = 0.641) (Table 10.2, Figure 10.3).

Table 10.2. Results of a Peto-Prentice test of significance of two watershed groups (A and B) with censored data for dissolved inorganic nitrogen and total phosphorus.

Watershed Group	n	n censored	Median	Mean	SD	p-value
Total Phosphorus (mg/L)						
A	46	25	0.016	0.017	0.019	0.906
B	43	29	0.007	0.016	0.018	
Total Nitrogen (mg/L)						
A	46	17	0.099	0.237	0.424	0.641
B	43	20	0.100	0.143	0.142	

Note: SD = standard deviation.

Discussion

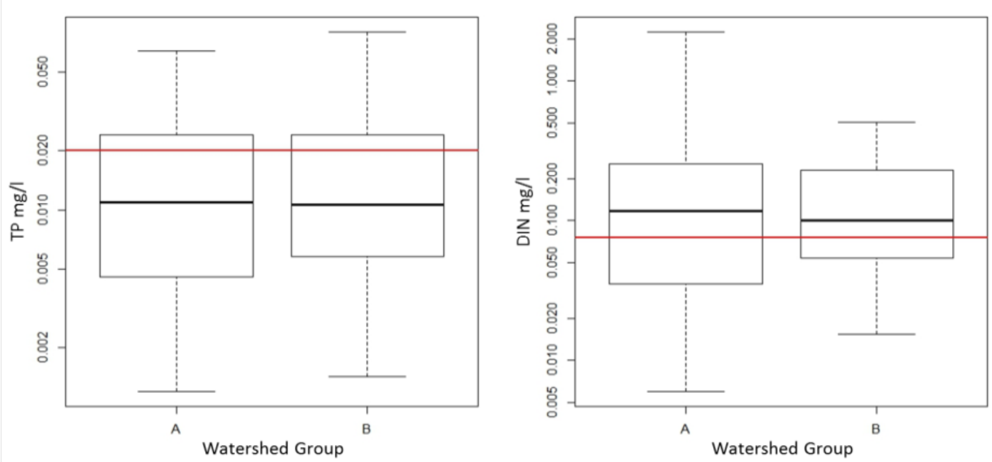


Figure 10.3. Boxplots showing distributions of total phosphorus (TP) and total dissolved inorganic nitrogen (DIN) between the two watershed groups (from k-clustering) of reference sites. Data below red vertical lines are censored and were extrapolated using Kaplan-Meier survival analysis.

GIS-based environmental gradients were used to explore classification by identifying the two groups of headwater streams that most strongly differed in environmental attributes (Figure 10.1). Starting with two groups was a somewhat arbitrary decision, but finer-scale classifications could have been explored if between-group differences in N or P concentrations were observed. No differences among Utah streams were identified, so further delineations were unnecessary.

Correspondence with Ecoregions

Trends revealed by the PCA axis loadings provided useful insights. The first PCA axis revealed two important environmental gradients: (1) those directly or indirectly associated with elevation (air temperature, elevation, precipitation), and (2) those associated with soils. Olson and Hawkins (2013) created watershed models to predict background TN and TP concentrations from watershed characteristics and found similar variables to be among the most important predictors of ambient nutrient concentrations. Many ecological attributes change when the location (upstream to downstream) along the stream changes, and such conditions are predicted to influence ecosystem processes such as the relative importance of autotrophic and heterotrophic processes (Vannote et al. 1980), which in turn can alter nutrient concentrations (Ensign and Doyle 2006). Differences in soil characteristics, especially nutrient content from the underlying lithology and erosion rates, are well known to contribute to nutrient inputs to streams. This observation is one reason for the selection of ecoregions as an *a priori* classification scheme when setting NNC (USEPA 2000).

Differences in nutrient concentrations were not found when comparing the two groups of streams that differed in these characteristics. It is possible that differences in nutrients among HUC-12 watersheds result from local attributes that are not easily captured with GIS-based data. It is also possible that important landscape-level attributes were missing in the analysis. Additional work will be needed to determine what characteristics, if any, lead to atypically high N or P within headwater watersheds. If such conditions are identified, they can be used to modify headwater NNC on a site-specific basis.

Relevance to Utah's Headwater Numeric Nutrient Criteria

These classification efforts focused on nutrients as the stressor, as opposed to ecological responses (Chapters 2–6), because ecological response data were not available for most of Utah's headwaters. An implicit assumption of this classification exercise is that ecological responses to nutrient enrichment are homogeneous between the two groups of physically distinct headwater streams. Many of the environmental gradients used to conduct the classification included direct or indirect measures of many physical characteristics that are known to affect the magnitude of ecological responses to nutrient enrichment, which suggests that this assumption may be valid. The earlier classification work included an evaluation of both nutrient concentrations and macroinvertebrate responses and did not identify distinct subclasses from either (Paul 2009). Nevertheless, it is possible that ecological responses to nutrient enrichment could systematically vary with physical characteristics that were not included in this analysis. As the headwater NNC are implemented, and response data are more broadly available, the validity of this assumption should be re-evaluated because NNC refined to appropriate subclasses should reduce errors in impairment determinations.

Data quality was another challenge encountered in this analysis. Many of the reference sites used for these classifications were sampled to support the development of biological assessment tools for DWQ. As a result, ambient nutrient concentrations used to characterize each site are often based on a single collection event. It is unclear whether this sample is comparable to the growing-season averages proposed for the headwater NNC. Furthermore, laboratory reporting limits were historically high, which resulted in many nondetects in the nutrient dataset. Organic N was lacking, so this analysis did not include TN, the nitrogen component of the headwater NNC. To a large extent, these issues have been resolved, and it may be worthwhile to collect additional data at reference sites in the future to better understand the naturally occurring ambient nutrient concentration among Utah’s headwater streams.

Perhaps the most important conclusion from these investigations is that headwater streams do not appear to require further classification for purposes of NNC development. This conclusion is consistent with previous efforts to classify all streams statewide with macroinvertebrate responses (Paul 2009). This result was contrary to expectations, particularly given climatic differences between the northern and southern regions in Utah, which are known to influence distributions of flora and fauna in the state (Omernik 1987). These analyses suggest that, at least in the context of ambient stream nutrients, natural environmental gradients associated with differences between mountains and valleys (e.g., temperature, gradient, size) may be more important than those associated with climate. Alternatively, it is possible that differences in nutrient concentrations among headwater streams are determined by local characteristics that are difficult to capture with the landscape-scale environmental characteristics available from GIS data sources.

Chapter 11

FREQUENCY DISTRIBUTION METHODS: AN EXPLORATION OF AMBIENT NUTRIENT CONCENTRATIONS IN HEADWATER STREAMS

Key Points

Historical data collected from headwater streams was compiled to evaluate the distribution of nitrogen and phosphorus among all headwater streams and those previously determined to be in reference condition.

Among reference sites the average total phosphorus (TP) concentration was 0.020 mg-P/L and the average total nitrogen (TN) concentration was 0.24 mg-N/L.

Nutrient concentrations were higher, although still relatively low, with an average TP of 0.030 mg-P/L and TN of 0.33 mg-N/L.

The range of reference site percentiles (75^{th}- 95^{th}) sometimes used to benchmark numeric nutrient criteria ranged from 0.027 – 0.053 mg-P/L for TP, and 0.29 – 0.61 mg-N/L for TN.

Modeled background nutrient concentrations for the majority of headwater sites ranged from 0.02 – 0.04 mg/L for TP and 0.2-0.3 mg/L for TN, which suggests that estimates derived from percentiles were reasonable estimates of naturally-occurring conditions.

Introduction

From an analytical perspective, frequency distribution methods (FDMs) are the most straightforward way to derive numeric nutrient criteria (NNC) (for details and applications see: Hawkins et al. 2010, Herlihy et al. 2008, Paulsen et al. 2008). FDMs derive NNC from the distribution of ambient nutrient concentrations in a region of interest, which can often be compiled from existing water quality databases. Most often, but not always, FDMs are restricted to data obtained from reference sites, which are assumed to be reflective of natural, background conditions (Hawkins et al. 2010, Stoddard et al. 2006). NNC are then defined by predetermined percentiles of ambient nutrient concentrations. The specific percentiles used to establish criteria are somewhat arbitrary, but historically the U.S. Environmental Protection Agency (USEPA) has recommended setting criteria at the 75th percentile of reference site nutrient concentrations, or the 25th percentile of all streams (USEPA 2000). This recommendation has received considerable criticism as being overly conservative because, among other things, it would result in the classification of ~25% of reference sites—streams that are, by definition, in the least degraded condition possible—as impaired (not meeting aquatic life uses). Another criticism of FDM methods is that they do not examine effects on aquatic life uses, and therefore assume that any increase in nutrients above background concentrations is deleterious to aquatic life, implying that streams lack assimilative capacity for nutrient enrichment.

USEPA has released guidance for using ambient nutrient concentrations among all streams in a region of interest to derive NNC (USEPA 2000), but criteria have not been promulgated using this approach. Nevertheless, the distribution of ambient nutrients obtained from all sites in a region of interest from both reference and nonreference streams can inform NNC development. For example, the range of nutrient concentrations among all streams can help managers understand the extent of impairments that are likely to occur if NNC were adopted, which can help them plan for future resource demands. Division of Water Quality (DWQ) has an interest in the number of new impairments that may be identified through the adoption of the proposed headwater NNC. Similarly, the extent of nutrient enrichment can be estimated in different regions of interest, which can help with watershed management because resources can be focused in areas where nutrient-related problems are of greater concern. Perhaps most importantly, many states, including Utah, require an evaluation of the economic impact of proposed rules, and statewide estimates of the number of impairments can help DWQ make more accurate estimates for these disclosures when the proposed NNC go through the rulemaking process.

This chapter summarizes ambient nutrient concentrations obtained from headwater streams throughout Utah over a nine-year period of record. Total nitrogen (TN) and total phosphorus (TP) concentrations collected from all headwater streams and from headwater reference sites are evaluated. Reference site distributions are compared with percentiles that have been used to derive NNC. Ambient nutrient concentrations collected from all sites are reviewed to evaluate the extent to which TN or TP concentrations vary among Utah's major watersheds and to estimate the number of impairments that could result from the proposed headwater NNC.

Methods

Data Compilation

Nutrient data were compiled for all headwater streams where one or more samples were collected over the nine-year period of record (2003–2012). The dataset was constrained to samples collected during the growing season, defined as June through September. Data were further screened to ensure that they did not include QA/QC samples, nor samples where field or laboratory notes suggested a potential concern with sample contamination or highly atypical conditions at the time of sample collection.

In several cases, > 1 sample was collected at the same site on the same day, in which case one sample for each day was randomly selected. Application of these screens to all potential water quality records resulted in ~6,800 records suitable for the evaluation of ambient nutrient concentrations among Utah's headwater streams.

All told, 494 unique headwater sample locations had available data for TP, and 448 sample locations had N data. Each site was sampled, on average, 6 times for the N dataset and 8 times for the P dataset during the growing seasons over the nine-year period of record. The number of growing-season samples collected at each site was highly variable, ranging from 1 to 43 collection events. Such unequal sample sizes are not uncommon for evaluations of nutrient concentration over broad spatial and temporal scales, but they can bias subsequent data exploration results. One concern with this dataset was to have geographically well-dispersed samples because these data will be used to inform NNC that will be applied statewide. Fortunately, nutrient samples were roughly equally distributed among Utah's major water management basins (Figure 11.1). The only exception was in the Colorado River basins where fewer samples were collected, which is likely a consequence of the fact that these watersheds are in drier areas of the state.

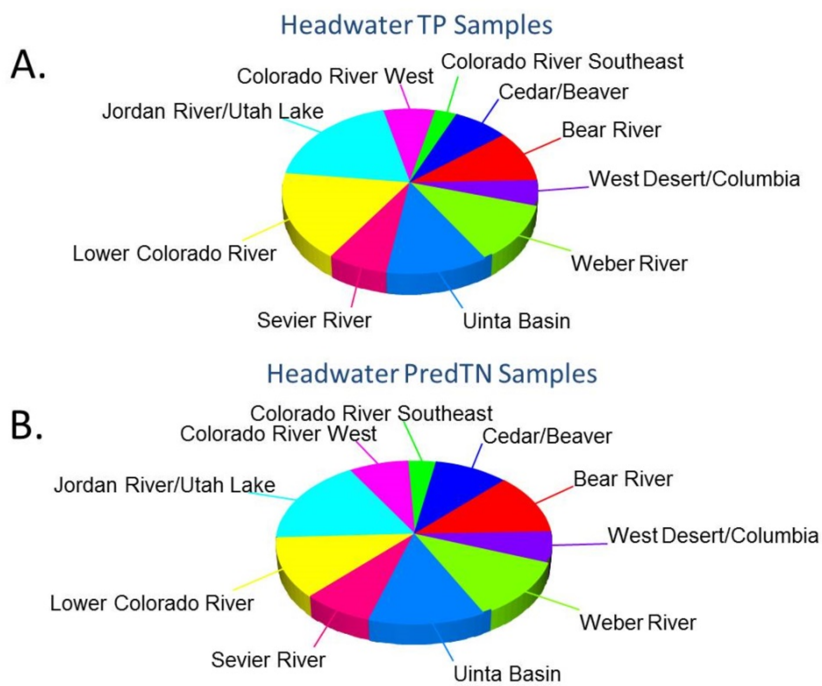


Figure 11.1. Distribution of total phosphorus (TP, Panel A, N = 3861) and predicted total nitrogen (PredTN, Panel B, N = 2947) samples among Utah's major watersheds (2003–2012).

Nutrient Concentrations at Headwater Reference Sites

The vast majority of Utah's headwater streams are located on public lands upstream of U.S. Forest Service (USFS) boundaries. These watersheds are generally well-managed and in reasonably good condition relative to downstream locations where the majority of Utah's population resides. As a result, many of the evaluated headwater sites may actually be in reference or minimally degraded condition. However, *post hoc* reference site determinations could not be made for most sites. Therefore, this study of reference sites was limited to those that DWQ had previously screened for upstream sources of human-caused stress and then evaluated to determine their physical and biological condition; this left 45 streams that could be used to evaluate background nutrient concentrations.

Estimation of Total Nitrogen

One limitation of historic water quality records was a paucity of TN data. Historically, only ammonia and nitrate-nitrite data were routinely collected by DWQ and cooperating agencies (i.e., USFS, Bureau of Land Management), which means that data on organic N was missing. Because of this data gap, TN could be calculated only for the most recent collection events. This issue had to be addressed to avoid exacerbating the differences in unequal sample sizes between the TN and TP datasets. This data limitation was identified several years ago, and ambient monitoring samples are now routinely analyzed for TN.

Most of the primary production within Utah's headwater streams is benthic; therefore, water column organic N concentrations should be relatively low and consistent among headwater streams. For this study, 193 headwater TN samples were available and were used to quantify a linear relationship between TN and total inorganic N (TIN; ammonia and nitrate-nitrite) that was remarkably consistent (Figure 11.2, $p < 0.001$, $r^2 = 0.92$). This relationship was used to predict TN (predTN) for samples from TIN for all samples where TN was unavailable.

Calculation of Growing-Season Average Ambient Nutrient Concentrations

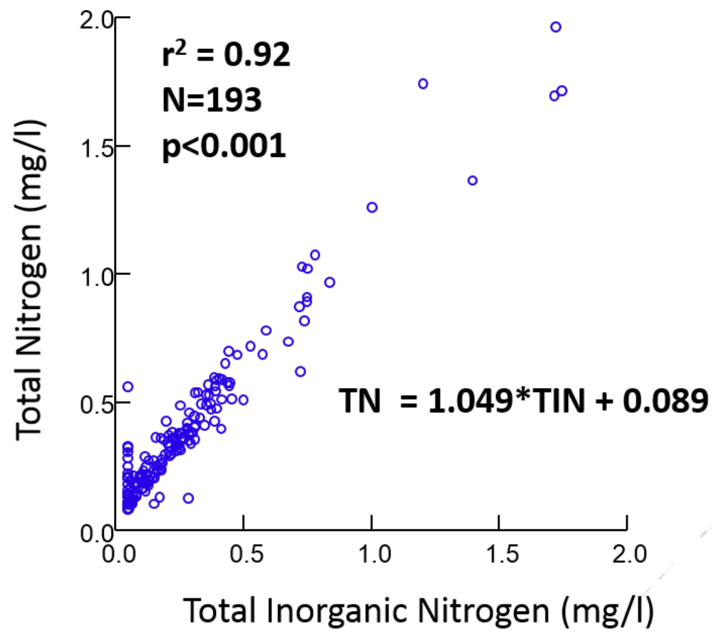


Figure 11.2. Total nitrogen as a function of total inorganic nitrogen (n = 193).

Given the unequal sample sizes among sites, it was necessary to summarize the data in a way that would neither over- nor under-represent the weight of a single site in distributions of ambient nutrient concentrations. At sites with > 1 sample, the average nutrient concentration among all growing-season samples was calculated. The use of growing-season averages normalized among-site differences and the number of collection events, which had the added advantage of expressing nutrient concentrations in a form similar to the proposed NNC. In circumstances where a single sample was available at a site, it was assumed to reflect the summertime average. While this assumption may not be appropriate for the evaluation of a single sample location, it was sufficient for this broad-scale evaluation with a large number of samples because any bias is equally likely to over- or underestimate average nutrient concentrations. These seasonal site averages were used for subsequent percentile calculations and associated distributions. For purposes of these analyses, values for all nondetects were half of reporting limits, a decision which should not alter the percentiles obtained from these distributions, provided that percentile concentration is above the laboratory reporting limit.

Results

Reference Site Nutrient Concentrations

Total Phosphorus

In all, 165 summertime TP samples were collected from 45 unique reference sites were evaluated (Table 11.1). Each reference site was sampled from 1 to 24 times over the nine-year record included in this review. The mean growing-season average TP among these sites was 0.020 mg-P/L, although average concentrations were considerably variable among reference streams (Figure 11.3, Panel B). FDM percentiles commonly used to support NNC (75^{th}–95^{th}) suggest that TP criteria using this approach would fall between 0.027 and 0.053 mg-P/L (Table 11.2).

Table 11.1 Summary statistics for the growing-season average of total phosphorus and predicted total nitrogen among headwater reference streams.

	Total Phosphorus (mg/L)	Predicted Total Nitrogen (mg/L)
Number of sites	45	42
Minimum	0.008	0.15
Maximum	0.073	0.66
Median	0.010	0.18
Arithmetic mean	0.020	0.24
Geometric mean	0.016	0.22
Standard deviation	0.016	0.12
Coefficient of variation	0.79	0.51

Table 11.2. Reference sites, common percentiles used for numeric criteria development based on growing season average ambient concentrations.

Percentile	Total Phosphorus (mg/L)	Predicted Total Nitrogen (mg/L)
75^{th}	0.027	0.29
90^{th}	0.037	0.38
95^{th}	0.053	0.61

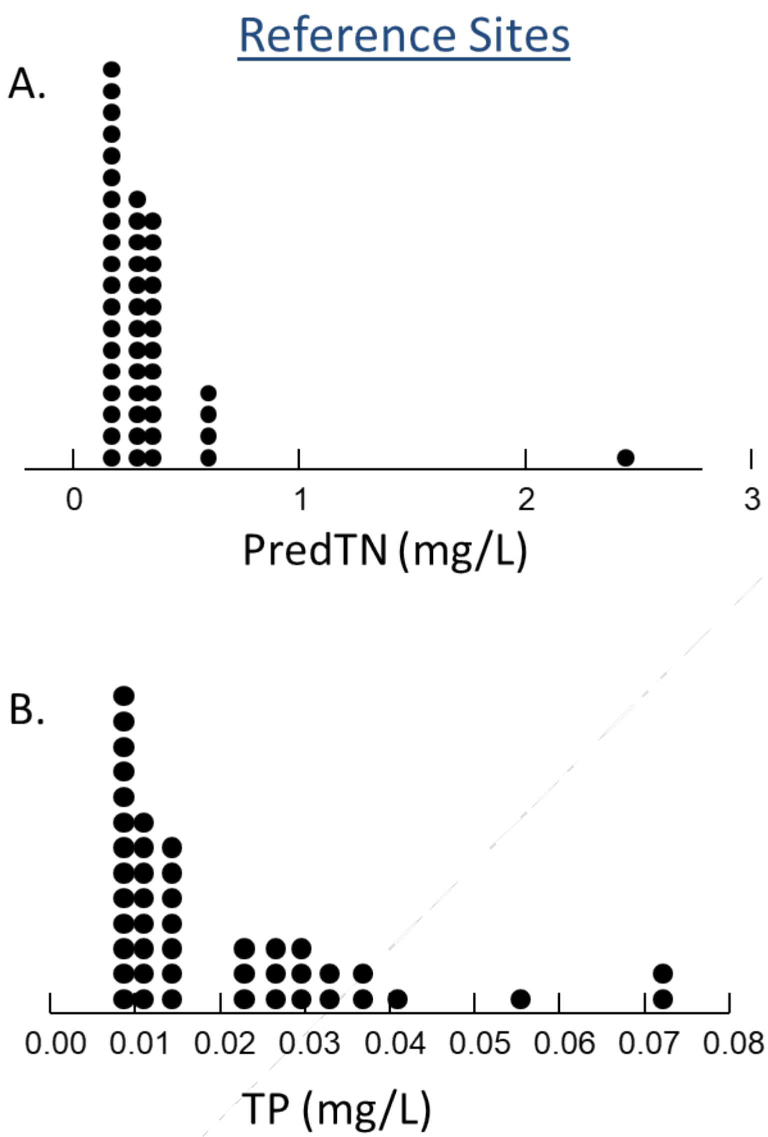


Figure 11.3. Distribution of the average summertime nutrient concentrations among reference streams for predicted total nitrogen (PredTN, Panel A, n = 43) and total phosphors (Panel B, n = 45).

Total Nitrogen

In all, 131 summertime predTN results were obtained from 43 unique reference sites (Table 11.1). Sites were sampled from 1 to 15 times over the period of record. On average, PredTN at these headwater reference sites was 0.24 mg/L, although average concentrations were considerably variable among reference streams (Figure 11.3, Panel A). The 75th to 95th percentiles among these reference sites ranged from 0.29 to 0.61 mg/l (Table 11.2).

Nutrient Concentrations among All Headwater Streams

Total Phosphorus

In all, 494 sites were sampled from 1–43 times during summertime months over the nine years that were evaluated, generating 3,861 headwater TP samples. Among all headwater sites, the mean TP was ~0.02–0.03 mg-P/L (depending on the measure of central tendency used to make this determination), which roughly corresponds to the 75th percentile of headwater reference streams (Table 11.3). While summertime TP concentrations were generally low, a few fairly high values (maximum = 0.471 mg-P/L) were recorded that would be considered impaired using the proposed NNC.

Table 11.3. Distribution of total phosphorus and predicted total nitrogen among all headwater streams.

	Total Phosphorus (mg/L)	Predicted Total Nitrogen (mg/L)
Number of cases	494	448
Minimum	0.006	0.09
Maximum	0.471	2.45
Median	0.019	0.24
Arithmetic mean	0.030	0.33
Geometric mean	0.021	0.27
Standard deviation	0.036	0.26
Coefficient of variation	1.194	0.78
75th percentile	0.038	0.38
90th percentile	0.058	0.56
95th percentile	0.083	0.88

Total Nitrogen

Patterns for summertime PredTN among all headwater streams were similar to those observed for TP. Fewer sites (N = 448) and samples (n = 2,943) were available for these analyses, but the available samples were distributed roughly equally among ecoregions. On average, summertime PredTN was ~0.3 mg/L among all headwater streams (Table 11.3). Interestingly, as with TP, this average concentration also roughly corresponds to the 75th percentile of headwater reference sites. While headwater PredTN was low for most summertime samples, a handful of samples were atypically high (maximum = 2.5 mg/L).

Discussion

Utah is the second driest state in the United States and the vast majority of the water that the state receives originates in headwater regions as snow. The critical importance of headwater resources has long been acknowledged by DWQ and other management agencies that have collectively implemented many different regulation and management programs intended to protect this important resource. From the perspective of nutrient pollution these efforts appear to have been successful because only ~25% of the streams evaluated exceeded the lower proposed NNC thresholds.

Frequency Distribution Approaches for Setting Numeric Nutrient Criteria

Traditionally, USEPA guidance has recommended setting NNC from the 75th percentile of reference site ambient nutrient concentrations or, in circumstances where reference site data are unavailable, setting them at the 25th percentile of all sites (USEPA 2000). This recommendation for the use of all sites assumed that both methods would converge on similar nutrient concentrations; however, several recent studies, including this one, have found that this assumption may not be valid. The 75th percentile of reference sites for TP was 0.027 mg/L, whereas the 25th percentile among all headwater streams was 0.010 mg/L. Similarly, for TN the 75th percentile of reference sites was 0.295 mg/L, whereas the 25th percentile among all streams was 0.168 mg/L. Suplee and colleagues (2007) found that the 75th percentile of nutrient concentrations in reference sites corresponded to the 4th-97th percentile of the general population of Montana streams, depending on the ecoregion. Rohm and colleagues (2002) found that the 75th percentile corresponded to the 40th-50th percentile of all reference streams. Evans-White and colleagues (2014) reviewed several studies and found a similar pattern. The results of these studies suggest that setting NNC from the 25th percentile of all sites is often more conservative than basing TN and TP thresholds on the 75th percentile of reference sites. The relatively large percentage of reference sites that are DWQ's best estimate of naturally occurring ambient nutrient concentrations that would exceed NNC if NNC were based on the 25th percentile of all sites suggests that these criteria would be overly conservative.

While the distribution of ambient nutrients observed at reference sites is informative, it should not be the only line of evidence used to support NNC. For instance, departure from reference condition is not necessarily the same thing as degradation of an aquatic life use. In some cases, small increases in nutrient concentrations can be beneficial to measures of biological integrity. Fisheries biologists have a long history of augmenting nutrients in highly oliogotrophic ecosystems to improve fish abundance and biomass (Johnston et al. 1990, Stockner et al. 1978, Stockner and Macisaac 1996). Eventually, as nutrient enrichment continues, these positive responses transition to deleterious responses (such as those discussed throughout this report). When deleterious changes begin, NNC specifications are warranted. An additional complication is that the specific nutrient concentration associated with a switch to deleterious responses differs from one stream to the next because physical conditions (e.g., channel shading, slope, water temperature) alter the susceptibility of a stream to nutrient enrichment. None of these limitations negate the use of distributional methods for setting NNC, but they do suggest that these methods should be used in concert with other lines of evidence.

Considerations for the Derivation of Headwater Numeric Nutrient Criteria

One potential use of reference site FDMs is the estimation of nutrient concentrations that reflect the point where human-caused enrichment is just starting to occur, because this could be used to identify nutrient concentrations above which ecological responses warrant further evaluation. The 90th percentile nutrient concentrations among reference sites (TN = 0.382 mg-N/L, TP = 0.037 mg-P/L) could be assumed to be the best definition of the upper end of naturally occurring ambient nutrient concentrations. To assess whether this was a reasonable estimate of background conditions, this measure was compared to an alternative approach that uses reference site nutrients and models to make site-specific predictions of background TN and TP from watershed characteristics such as local lithology, precipitation, and soil structure (Olson and Hawkins 2013). While spatially variable, the proposed lower thresholds (TN = 0.40mg/L and TP = 0.035 mg/L) align with the mid-to-upper range of model predictions of background nutrient concentrations using this technique (Figure 11.4).

USEPA guidance recommends basing NNC on the 75^{th} percentile of reference sites; the stressor-response models used in this technical support document suggest this approach may be overly conservative because this concentration falls below the nutrient concentrations associated with the lower ecological threshold for almost every ecological response evaluated (Chapters 2–7). Others have made similar observations. For instance, Heatherly (2014) compared criteria set at the 75^{th} percentile of reference sites with criteria that considered ecological responses and found the latter 5–6 times lower. Supplee and colleagues (2007) summarized several studies that reported nutrient concentrations thought to be harmful to biota in mountain streams and noted that concentrations of threat to aquatic life uses were equivalent to ~86^{th} percentile of reference site nutrient concentrations in Montana. NNC have also recently been promulgated—and approved—by USEPA that correspond to higher reference site percentiles, although USEPA has not issued guidance on what other percentiles would be considered to be appropriately protective of aquatic life uses.

Figure 11.4. Modeled predictions of background TP (left panel) and summertime TN (right model).

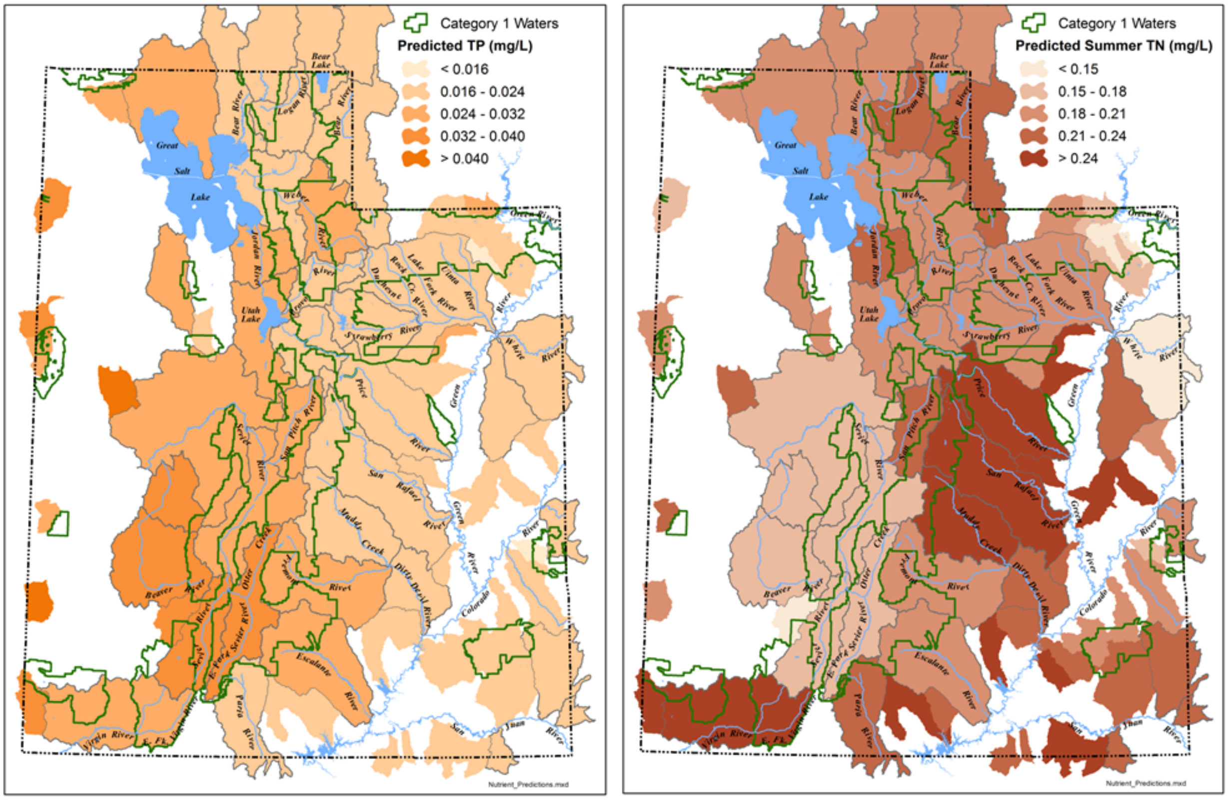


Figure 11.4 Backgroud TN and TP predicted as predicted from reference site emirical models.

SECTION 3

NUMERIC NUTRIENT CRITERIA FOR HEADWATER STREAMS

Chapter 12

NUMERIC NUTRIENT CRITERIA FOR HEADWATER STREAMS: MULTIPLE LINES OF EVIDENCE

Key Points

The evidence presented in previous chapters of this report were reviewed to generate a proposal for numeric nutrient criteria (NNC) that uses a combination of ambient nutrient concentrations and ecological responses t0 assess support of aquatic life uses.

The NNC proposes ranges of ambient nutrient concentrations: a lower range considered to be reflective of naturally occurring conditions, an upper range considered to be reflective of substantial enrichment and a middle range where responses are needed to determine whether or not aquatic life uses are supported.

The NNC proposes using gross primary production and filamentous algae cover to evaluate autotrophic responses to nutrient enrichment and ecosystem respiration to evaluate potential heterotrophic responses.

Introduction

The numeric nutrient criteria (NNC) for headwater streams proposed in this chapter are the culmination of a multi-year collaborative effort between Division of Water Quality (DWQ) and stakeholders. Leaders from state and federal agencies, nonprofit organizations, and the private sector—the *Nutrient Core Team*—were brought together to help develop and prioritize initiatives that DWQ could implement to reduce anthropogenic sources of nutrients to Utah's waterbodies. This group recommended prioritizing the development of regional NNC for headwater streams. The NNC are based on analyses developed and refined by a *Technical Review Team* of scientists.

The NNC for headwater streams presented in this chapter were developed using data and information presented in previous chapters of this technical support document (TSD) and scientific evidence published elsewhere. Headwater streams were assessed to ascertain whether they could be grouped into useful subclasses that would help minimize natural variation in ambient nutrient concentrations, but no empirical support for subclasses was found (Chapter 10). Frequency distribution methods (FDMs) were used to examine the distribution of ambient total nitrogen (TN) and total phosphorus (TP) concentrations among reference sites to help estimate naturally occurring nutrient concentrations (Chapter 11). Several stressor-response (S-R) models were developed to establish quantitative linkages between ambient nutrient concentrations and ecological attributes considered to be indicative of biological condition (Chapters 2–7). Where possible nutrient concentrations associated with the empirically derived S-R thresholds were compared to those associated with independently derived measures of support for aquatic life uses. Because nutrient enrichment also has the potential to degrade aquatic life uses, a survey that evaluated the extent to which excessive benthic algae biomass could alter decisions about whether or not recreation at a stream was desirable to Utah citizens was completed (Chapter 8).

The breadth of available evidence lends itself to the derivation of NNC using a Weight of Evidence (WoE) approach. WoE is a decision-making process that is often used in regulatory contexts because it allows simultaneous consideration of all the data and information available (Menzie et al. 1996, Smith and Tran 2010). Such approaches are scientifically defensible because they more accurately capture the breadth and complexity of complex ecological relationships than approaches that rely on a single response or investigation. However, distillation of multiple lines of evidence into a single value that can be used for regulatory purposes can be challenging, and professional judgment is often necessary to make such decisions.

This TSD takes the approach that any increase in nutrient concentrations over ambient conditions could be considered degradation of water quality—with the important caveat that minor excursions over ambient conditions may or may not be associated with harm to the use. This chapter presents a conceptual model where ambient nutrient concentrations could be used to place streams into one of three bins: (1) indistinguishable from reference condition; (2) some enrichment has likely occurred, but the evidence suggests that this may or may not be sufficient to harm aquatic life uses; and (3) human-caused enrichment is highly probable and likely extensive (broader spatial and temporal extent, Figure 12.1).

The proposed NNC place headwater streams into one of three bins (Figure 12.1, Table 12.1) based on the average ambient TN and TP concentrations during the period of algae growth through senescense. Streams are considered to be fully supporting of aquatic life uses if ambient concentrations of both TP and TN fall below the lower thresholds of 0.035 mg/L and 0.40 mg/L, respectively, unless ecological response data suggest that nutrient-related degradation has occurred. The proposed NNC also define upper thresholds of 0.080 mg TP/L and 0.80 mg TN/L; a stream with concentrations above these thresholds would be considered to be failing to support aquatic life uses, irrespective of the availability of ecological response information. Finally, the proposed NNC requires that streams between these thresholds have information on one of three ecological responses—gross primary production (GPP), ecosystem respiration (ER), or filamentous algae cover—to make an attainment determination. These proposed aquatic life use NNC will be combined with the recreational use values proposed in Chapter 8 for all of Utah's headwater streams.

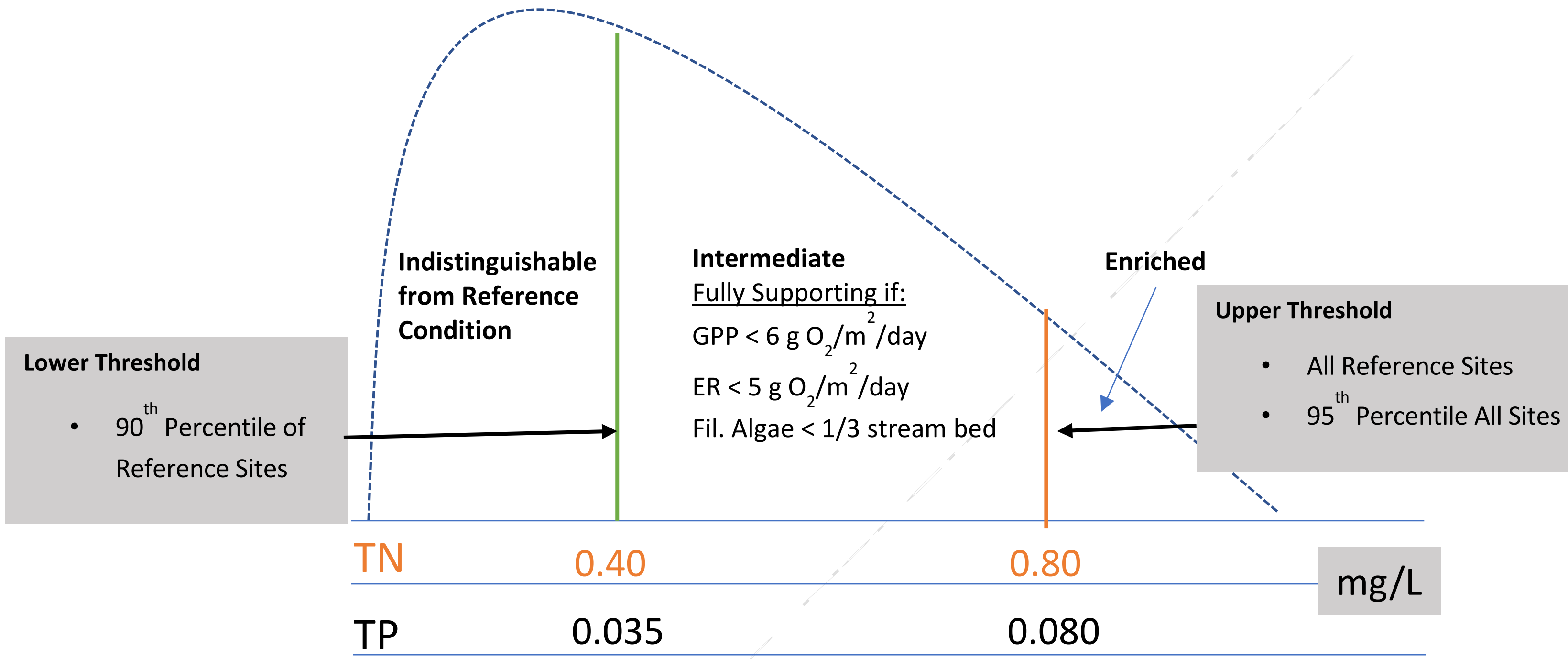


Figure 12.1. Hypothetical histogram of nutrient concentrations depicting proposed thresholds for total nitrogen and total phosphorus and ecological responses needed to interpret aquatic life use support at intermediate headwater streams.

Notes: ER = ecosystem respiration, GPP = gross primary production, TN = total nitrogen, and TP = total phosphorus.

Table 12.1. Numeric Nutrient Criteria and Associated Ecological Responses (Bioconfirmation Criteria) Proposed to Protect Aquatic Life Uses in Antidegradation Category 1 and 2 (UAC R317-2-12)[f] Headwater Perennial Streams

Low Nutrient Enrichment at Headwater Streams: No Ecological Responses			
Summertime Average Nutrients		Assessment Notes	
TN < 0.40[a,b]	TP < 0.035[a,b]	Fully supporting aquatic life uses if the average of ≥ 4 summertime samples for both TN and TP fall below the specified nutrient concentrations. However, it is not supporting aquatic life uses, cause unknown, if the ecological responses specified for moderate enrichment streams are exceeded. Sites with fewer samples, or those without TN and TP growing season averages, will not be assessed for nutrients.	
Moderate Nutrient Enrichment at Headwater Streams and Ecological Responses			
Summertime Average Nutrients		Ecological Response	Assessment Notes
TN 0.40– 0.80[a]	TP 0.035– 0.080[a]	Plant/Algal Growth[c] < 1/3 or more filamentous algae cover[d,e] OR GPP[c] of < 6 g O_2/m^2/day or ER[c] of < 5 g O_2/m^2/day	Headwater streams within this range of nutrient concentrations will be considered impaired (not supporting aquatic life uses) if <u>any</u> response exceeds defined thresholds. Streams <u>without response data</u> will be listed as having <u>insufficient data</u> and prioritized for additional monitoring if either TN or TP falls within the specified range.
High Nutrient Enrichment at Headwater Streams: No Ecological Responses[e]			
Summertime Average Nutrients		Assessment Notes	
TN > 0.80[a,b]	TP > 0.080[a,b]	Streams over these thresholds will initially be placed on Utah's Section 303(d) list as threatened. Threatened streams will be further evaluated using additional data such as nutrient responses, biological assessments, or nutrient-related water quality criteria (e.g., pH and DO) both locally and in downstream waters.	

<u>Notes</u>: Criteria are applicable during the period of algae growth through senescence unless more restrictive total maximum daily load (TMDL) targets have been established to ensure the attainment and maintenance of downstream waters. DO = dissolved oxygen, ER = ecosystem respiration, GPP = gross primary production, TN = total nitrogen in mg/L, and TP = total phosphorus in mg/L.
a. Seasonal average of ≥ 4 samples collected during the period of algae growth through senescence will not be exceeded. Sites will be assessed using the higher of TN and TP threshold classifications.
b. Response data, when available, will be used to assess aquatic life use support or as evidence for additional site-specific investigations to confirm impairment or derive and promulgate a site-specific exception to these criteria.
c. Daily whole stream metabolism obtained using open-channel methods. Daily values are not to be exceeded on any collection event.
d. Filamentous algae cover means patches of filamentous algae > 1 cm in length or mats > 1 mm thick. Not to be exceeded daily stream average, based on at least 3 transects perpendicular to stream flow and spatially dispersed along a reach of at least 50 meters.
e. Quantitative estimates are based on reach-scale averages with at least three measures from different habitat units (i.e., riffle, run) made with quantitative visual estimation methods.
f. Excluded waters identified in UAC R317-2-14, Footnotes for Table 2.14.7 and Table 2.14.8.

The remainder of this chapter begins with an explanation of the rationale used to define the proposed NNC thresholds for TN and TP. The selection of GPP, ER, and filamentous algae as ecological responses in the combined proposed NNC is explained, as is the rationale used to establish the proposed NNC nutrient and response thresholds. The chapter concludes with a thorough discussion of considerations that went into the development of the proposed NNC and additional evidence that supports the proposed NNC as scientifically defensible, ecologically meaningful, and appropriately protective of aquatic life uses. The chapter following this one provides the results of an independent investigation that was conducted to evaluate the proposed NNC.

Derivation of Numeric Nutrient Criteria Thresholds

Nutrient Thresholds

The proposed lower nutrient NNC thresholds (growing-season average not to be exceeded levels: TN = 0.40mg/L, TP = 0.035 mg/L) are intended to reflect concentrations at which (1) anthropogenic nutrient enrichment has almost certainly started to occur, but (2) the concentrations are below levels associated with biologically degraded conditions. With respect to the first consideration, excursions of the 90^{th} percentiles of reference site nutrient concentrations were assumed to be related, at least in part, to human-caused nutrient inputs. While the analysis of nutrient concentrations at reference sites was based on a moderate sample size, reference site FDM results published for similar streams suggest that these estimates of ambient nutrient concentrations are reasonable (Figure 12.1). For example, Evans-White and colleagues (2014) identified peer-reviewed publications that reported reference site nutrient concentrations collected at streams in U.S. Environmental Protection Agency (USEPA) Aggregate Nutrient Ecoregion II (Western Forested Mountains). At these streams, the 75^{th} percentile of reference sites ranged from 0.08–0.21 mg/L for TN and 0.003–0.020 mg/L for TP. The higher concentrations of these distributions are nearly identical to the observations presented here (TN = 0.27 mg/L, TP = 0.294 mg/L; Chapter 11).

To ensure that the proposed lower NNC thresholds were protective of aquatic life uses, the proposed lower NNC thresholds were compared with thresholds derived from S-R models (Figure 12.2). The results of numerous S-R models quantified relationships between ambient nutrient concentrations and numerous ecological responses (Table 12.2, Chapters 2–7). For most responses the S-R models were used to derive a lower threshold that could be used to distinguish between streams in good versus fair condition (based on the response) and an upper threshold to distinguish between streams in fair versus poor condition (Table 12.3). The lower threshold (the first inflection point in the S-R relationship) was interpreted as reflecting the initiation of a measurable departure from natural conditions, which may or may not be of sufficient magnitude to constitute a degradation of aquatic life uses. The upper threshold was interpreted as the point where changes to responses were considerable and were highly likely to be reflective of impaired conditions (adverse effects to aquatic life uses).

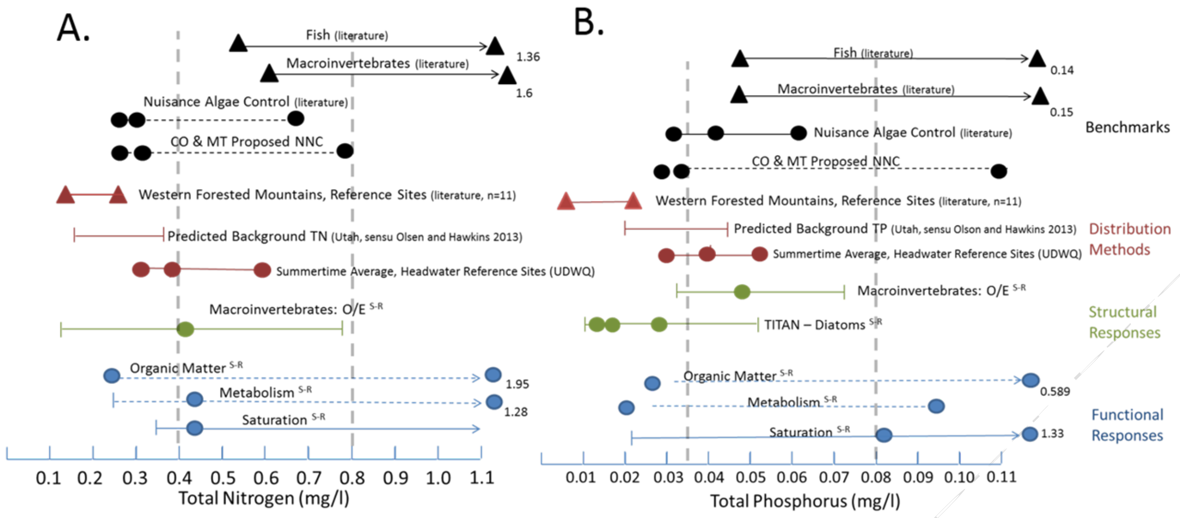


Figure 12.2. Numeric nutrient criteria thresholds derived from numerous sources for total nitrogen (panel A) and total phosphorus (panel B), along with the proposed numeric nutrient criteria for these nutrients presented in this technical support document.

Notes: Lines bracketed by triangles indicate the omission of numerous intermediate thresholds (dots). The graphics are colored to demarcate different categories of thresholds. Blue denotes functional responses. Green denotes structural responses (Division of Water Quality calculations). Red denotes thresholds derived using frequency distribution methods; the bottom red dots indicate the 75th, 90th, and 95th percentiles of the summertime average of Utah reference sites, the middle red line denotes background concentrations obtained from an empirical model that predicts background concentrations from natural environmental gradients, and the top red line denotes other distribution methods from reference site distributions in USEPA Nutrient Ecoregion II (Evans-White et al. 2014). Black denotes broad benchmarks for other proposed numeric criteria from USEPA Region 8 (the bottom black line) and values obtained from primary literature (the top three black lines; Evans-White et al. 2014). The vertical dotted lines are the proposed numeric nutrient criteria thresholds presented in this technical support document.

Table 12.2. Summary of indicators used to guide development of proposed headwater numeric nutrient criteria thresholds.

Functional Indicators	
Nutrient Saturation	These thresholds, derived from nutrient-diffusing substrates, quantify the concentration of total nitrogen and total phosphorus where, on average, additional nutrients did not cause an increase in algal growth. At these thresholds, other factors, such as light, substrate, or CO_2, cause additional algal growth.
Stream Metabolism: Gross Primary Production (GPP)	GPP measures the total amount of oxygen produced by photosynthesizing plants and algae each day (g O_2/m^2/day). Nutrient thresholds derived from GPP are the concentrations that were associated with streams with relatively low, moderate, and high rates of GPP. DWQ proposed to use this as a water quality criterion paired to nutrient criteria.
Stream Metabolism: Ecosystem Respiration (ER)	ER measures the oxygen consumed either through the processing (oxidation) of organic matter to CO_2 or by plant and algae growth (g O_2/m^2/day). Many stream organisms—including bacteria, protozoa, and fungi—rely on organic carbon as an energy source for cellular metabolism and growth, and these processes consume oxygen. DWQ proposes to use this metric as a water quality criterion paired with nutrient criteria.
Organic Matter Standing Stock	Organic matter standing stocks quantify, as g C/m^2 or ash free dry mass (AFDM)/m^2, the amount of organic matter, excluding larger particle and macrophytes, in stream reaches. This measure provides an estimate of the amount of material available to feed bacterial, protozoan, and fungal respiration. Nutrient thresholds were derived as the concentrations that, on average, distinguished among streams with relatively low, moderate, and high organic matter standing stocks.
Structural Indicators	
TITAN (nCPA): Diatoms	Threshold indicator taxon analysis (TITAN; Baker and King 2010) is a method that calculates indicator scores that capture the occurrence, abundance, and directionality of species' responses to stressors. The method then uses these indicator scores to determine, in this case, the nutrient concentrations (nonparametric change point analyses [nCPA]) associated with statistically significant changes in composition. As used in this analysis, TITAN scores capture biological changes associated with diatom taxa, a diverse assemblage of algae that are known to be sensitive to nutrients.
Macroinvertebrate Biological Assessments: O/E	DWQ currently uses macroinvertebrate-based river invertebrate prediction and classification system (RIVPACS) models to evaluate biological integrity and to determine whether streams are meeting their designated aquatic life uses. The output of the models, O/E, is a ratio of the number of macroinvertebrates that were actually observed at a site compared with the number of species predicted to occur in the absence of human-caused stressors. Nutrient thresholds were derived for concentrations that best distinguished between streams in degraded and nondegraded conditions.

Table 12.3. Nutrient thresholds derived for various ecological responses from stressor-response modeling efforts.

Ecological Responses	Total Nitrogen (mg/L)	Total Phosphorus (mg/L)
Functional Indicators		
Nutrient Limitation	0.42	0.080
Stream Metabolism		
Lower Threshold	0.24	0.026
Upper Threshold	1.28	0.090
Autochthonous Organic Matter Standing Stock		
Lower Threshold	0.24	0.026
Upper Threshold	1.95	0.590
Structural Indicators		
TITAN		
Sensitive Macroinvertebrates	0.18	0.011
Tolerant Macroinvertebrates	0.41	0.610
All Macroinvertebrates (nCPA)	0.41	0.015
All Diatom Taxa (nCPA)	--	0.045
Biological Assessments		
Macroinvertebrate O/E	0.43	0.045
ROC Thresholds O/E	0.32	0.030

Notes: nCPA = nonparametric change point analysis, O/E = the ratio between the number of observed species and the number of species expected, ROC = receiver operating characteristic, and TITAN = total indicator taxon analysis.

Lower nutrient thresholds were based on two criteria. First, they should fall below the nutrient concentrations associated with the fair–good response thresholds for all but the most sensitive of nutrient responses. Second, the proposed lower NNC thresholds for TN and TP should also fall below the concentrations associated with <u>all</u> the fair–poor response thresholds because this was interpreted to be reflective of appreciable degradation to aquatic life uses. Candidate NNC thresholds were also evaluated against nutrient concentrations associated with other independently derived measures of aquatic life support (e.g., pH, dissolved oxygen [DO], the ratio of the number of observed species compared to the number of expected species [O/E]) to provide additional conformation that the lower threshold was protective of aquatic life uses using these independely derived condition indicators. The upper nutrient thresholds (growing-season average not to be exceeded levels: TN = 0.80 mg/L, TP =0.080 mg/L) are intended to reflect conditions where human-caused enrichment of streams has almost certainly occurred and, in many cases, is likely to be appreciable. Nutrient inputs should be reduced as quickly as possible for these streams to avoid long-term nutrient increases in enrichment to downstream waterbodies. The proposed upper threshold is above the highest growing-season average TN and TP concentration observed among reference sites (Chapter 11). This threshold is also above almost all of the S-R thresholds that were derived to distinguish between streams in good and fair condition, such that excursions above the upper nutrient threshold are likely to result in impairments to aquatic life uses, eventually, particulary if multiple reaches are evaluated, which would be an integral part of follow-up investigations to these impairments.

Ecological Responses

The wide range of indicators evaluated was informative and allowed exploration of a variety of potential responses to nutrient enrichment. The separate evaluation of ecological responses provides independent confirmation that the proposed NNC thresholds are broadly protective of aquatic life uses. Although continued monitoring of all ecological responses would provide more data, it is not tenable due to limited resources and logistical constraints; therefore, a subset of responses was selected for inclusion in the proposed NNC. Indicators of alterations to both autotrophic and heterotrophic production were selected because together they represent the two principal causal paths that can lead to the degradation of aquatic life uses. For reasons discussed further below, GPP and filamentous algae cover were selected as indicators of autotrophic conditions, and ER was selected as the indicator of heterotrophic condition. DO and pH criteria and O/E assessment thresholds were considered for inclusion in the combined NNC because excursions of each have been linked to nutrient enrichment. These parameters were not included in the proposed NNC and will be monitored independently because excursions of each can be caused by factors other than nutrient enrichment. The thresholds for GPP (6 g O_2/m^2/day) and ER (5 g O_2/m^2/day) are based on the stream metabolism S-R thresholds that best distinguished streams in good or fair ecological condition (the lower response thresholds identified by the S-R models). A filamentous algae cover response (filamentous algae not to exceed one-third of the wetted stream bed) was added to the proposed NNC after the S-R investigations revealed the need for a measure of primary production potentially harmful to stream biota that could be routinely collected alongside ambient nutrient samples. Like nutrient thresholds, each ecological response should be interpreted as a not-to-be-exceeded threshold of the proposed NNC.

Discussion

Regional NNC require that nutrients—and in some cases ecological responses—be distilled into specific numbers that can be used for regulatory purposes. One challenge in the selection of NNC thresholds was the evaluation of the relative strength of each line of evidence. Indicator strengths were evaluated using factors such as the strength of the underlying analysis, the nature of the relationship between nutrients and responses (e.g., is the effect direct or indirect), and the relative applicability of responses as measures of biological integrity (the regulatory objective). While general rules such as these can be consistently applied to all lines of evidence qualitatively, they cannot be directly compared quantitatively because they are measured on different scales and are not equally reflective of the indicators' relative importance to overall support of aquatic life uses.

In weighing the various lines of evidence, relevance to the Clean Water Act (CWA) obligation to be protective of aquatic life uses was among the most important considerations. On one hand, each of the responses evaluated could be considered to be of equal importance because each measures a unique aspect of biological integrity, which itself is interpreted as a holistic description of stream conditions. On the other hand, the responses differ with respect to their relevance in the subsequent guidance and regulations that have been employed to meet the CWA objective of protecting and maintaining the biological integrity of the nation's waters. For instance, USEPA (1985) states that the degradation of aquatic life means, "*Unacceptable long-term or short-term effects on…fish and invertebrate species in rivers and streams.*" The focus on fish and invertebrates has been integral to the development of biological assessment programs developed to measure support of aquatic life uses. For these reasons, slightly greater weight was placed on thresholds associated with Utah's O/E scores, which is currently how the biological integrity of streams is quantified. This does not mean that the other responses are unimportant. The responses included in the functional S-R models were critical to development of the proposed NNC because many are more directly tied to nutrient enrichment than the O/E scores. Therefore, the lower nutrient thresholds of the proposed NNC are generally lower than those associated with responses that distinguished between good and fair condition classes for all but the most sensitive responses evaluated (e.g., sensitive diatoms). All the proposed lower nutrient thresholds are also lower than the thresholds that distinguish between fair and poor condition classes for the evaluated ecological response metrics.

DWQ's monitoring and assessment program has practical constraints that were considered when developing the proposed NNC, which were defined to facilitate routine data collection of the parameters necessary for making decisions about support of aquatic life in headwater streams. This means the protection of aquatic life uses will not be impeded by a lack of sufficient data or information necessary to make regulatory decisions.

Considerations for Duration and Frequency Components of Numeric Nutrient Criteria

Many of the frequency and duration components of the proposed NNC were based on practical constraints. For example, given the high temporal variation of ambient nutrient concentrations (Arheimer and Liden 2000, Stutter et al. 2008), the requirement to calculate growing-season average nutrient concentrations on a minimum of four or more samples is lower than what ideal circumstances would dictate. However, the regulatory obligation to monitor and assess waterbodies statewide means that it is impractical to collect water chemistry samples more frequently than monthly. One danger of smaller sample sizes is that an atypically high nutrient concentration could cause a stream to exceed NNC thresholds due to the larger weight of each sample in growing-season average calculations. This could occur if nutrient samples were collected immediately following a storm event, when ambient concentrations can be higher than they are under base flow conditions (Verheyen et al. 2015). However, in most cases the additional NNC requirement for ecological response information would prevent these circumstances from leading to an erroneous conclusion of impairment.

For both nutrient concentrations and responses, "not to be exceeded" NNC recurrence intervals are warranted based on practical constraints. DWQ currently rotates monitoring among six major watersheds yearly, which means that requiring data collected in more than one season could result in impairments being unaddressed for 6 or more years. TN and TP NNC components are assumed to be integrative because they are based on summertime averages, as are measures of ecological responses. The decision to limit average nutrient concentrations to the growing season was based on both practical and scientific recommendations; the growing season is when autotrophic and heterotrophic production are generally highest in temperate streams (Francoeur et al. 1999) and are most likely to affect aquatic life use attainment. The growing season is defined as June–September based on published reviews for streams in the intermountain west (Suplee et al. 2007). Ending the season in September is also practical because this corresponds to the end of Utah's water year, which demarcates the beginning and end of DWQ's yearly monitoring plans.

Autotrophic Responses

Evidence from other investigations generally suggests that the proposed NNC nutrient thresholds are protective against excessive primary production in streams. Biggs (2000) recommended that dissolved inorganic N and soluble reactive P remain below 0.019 and 0.002 mg/L, respectively, to avoid nuisance algae growth (200 mg/m^2 chlorophyll-*a* [chl-*a*] for a 50-day accrual). *Cladophora*, a filamentous algae that sometimes leads to nuisance algae growth in Utah streams, has a higher likelihood of reaching nuisance levels when TN exceeds 0.6–1 mg/L or TP exceeds 0.02–0.04 mg/L (Dodds et al. 1997, Stevenson et al. 2006), although the extent to which nuisance levels are attained depends on the frequency and magnitude of flooding events (Freeman 1986) and other stream characteristics that limit algal growth (e.g., high canopy cover, high water velocity, unstable substrate).

Filamentous Algae Cover

While limited, a handful of studies have linked filamentous algae cover to nutrient enrichment; this is supported by observations of Utah's headwater streams. Stevenson and colleagues (2006) found the probability of getting a filamentous algal cover of 20–40% increased when TP was > 0.03 mg/L or TN was > 1 mg/L in Midwest United States streams considered susceptible to filamentous algae growth. This study also noted that filamentous algae was absent at many streams with high nutrients. Others have noted that whether filamentous algae cover reaches levels of potential concern depends on other stream characteristics, such as canopy cover, stream temperature, stream size, and hydrology (Busse et al. 2006, Dodds and Oakes 2004, Riseng et al. 2004). As a result, the amount of filamentous algae cover within a given stream can vary both seasonally and year-to-year (Dodds and Gudder 1992, Francoeur et al. 1999). These studies largely demonstrate that high concentrations of nutrients are necessary to sustain filamentous algae at levels high enough to cause degradation of aquatic life uses and that local site conditions are important determinants of whether this occurs. This is similar to other water quality problems that only manifest when conditions are most limiting. It also means that a definitive impairment conclusion can be made from an observation of high levels of filamentous algae cover, while the opposite conclusion—that there are no nutrient-related threats to aquatic life uses—may require several seasons of sampling to state definitively.

Much of the discussion surrounding filamentous algae in the scientific literature has focused on declines in aesthetics and increased nuisance conditions that ultimately degrade recreational uses. One possible reason is that recreational uses may be more sensitive to filamentous algae growth than aquatic life uses because the most severe adverse effects to aquatic life are associated with a dominance of benthic production by this growth form. Once filamentous algae dominates the streambed, numerous detrimental effects on aquatic life have been documented including: impairment of fish spawning success, habitat degradation, reduction in fish feeding activity, reductions in fish populations, increases to pH and subsequent increases in ammonia toxicity, and the potential for hypoxia during senescence of large amounts of algal biomass. DWQ is concerned that prolonged enrichment, particularly when combined with stabilization of flow via water diversions or impoundments, can cause streams to shift from diatom-based food webs to a state where the majority (i.e., >50%) of the stream bed is covered in filamentous algae. Such shifts have been observed in Utah streams and these prolonged periods of extensive filamentous algae blooms are considered to be an important indicator of eutrophication with a multitude of potential adverse effects to stream biota.

Biggs (2000) provides the most extensive review of filamentous algae cover as an indicator of stream nutrient enrichment. Much of this review focused on nuisance (recreational) adverse effects, but the review also contains considerable information about potential adverse effects to aquatic life, which is summarized below.

Extensive and long-lived filamentous algae blooms can degrade stream habitat in several ways. First, structure of these algae mats are known to trap suspended sediment, which can fill interstitial spaces in cobble-bedded streams. This diminishes the quantity and quality of macroinvertebrate habitat because they frequently reside in these interstitial spaces. Extensive algal mats also slow the flow of water, which can increase stream temperature in areas exposed to sunlight. Increases in water temperature can have adverse metabolic effects on macroinvertebrates because these organisms are ectotherms, with body temperatures dictated by their surrounding environments. Higher stream temperatures also decrease DO saturation, which can be detrimental to stream biota.

Another adverse effect of a regime shift to a condition where filamentous algae blooms are extensive (i.e., >50% cover) is alteration of stream food webs. For most stream macroinvertebrates, diatoms are a preferable food resource because they are higher in lipids and more easily digestible than filamentous algae. Some algal piercing taxa are specialized in feeding on filamentous algae and they will replace less specialized taxa when algal mats are present. However, these taxa are also less likely to actively drift, which can be detrimental to the feeding efficiency of trout and other fish. Decreases of trout abundance in streams dominated by filamentous algae have been documented and this may be due, in part, to shifts in the taxonomic composition of lower trophic levels. In mesotrophic streams, the loss of actively drifting stream insects is less severe because algal mats are more short-lived, allowing time for recolonization of these insects. Another impact of excessive filamentous algae growth is a reduction in the grazing efficiency of macroinvertebrates. Top down control of algal biomass through grazing has been well-documented in streams where stream autotrophs consist primarily of diatoms (the natural condition of Utah headwater streams), but these biotic controls on excessive benthic production are less effective once filamentous algae are the dominant forms of these ecosystems.

Extensive mats of filamentous algae also exacerbate other water quality parameters that have been demonstrated to be deleterious to aquatic life uses. Higher rates of primary production are associated with increases in pH, which has the potential to negatively affect the condition of stream fish and indirectly affects all aquatic biota by increasing the toxicity of ammonia that may be present at higher concentrations in nutrient enriched streams. The resulting relatively higher biomass of autotrophs can cause appreciable increases in the diel flux of DO, which has been attributed to biological degradation. The accrual of autochthonous carbon in streams with extensive filamentous algae mats can be considerable, especially when nutrient concentrations are high because episodic sloughing decreases and filamentous algae mats become longer and thicker due to increases in nutrient diffusion into the mats. In Utah, decreases in stream temperatures will ultimately result in algal senescence, which allows the carbon tied up in these mats more available to stream heterotrophs. Subsequently, this increases ER, which has the potential to cause hypoxic conditions (Suplee et al. 2019). The potential for delayed hypoxia highlights another advantage of filamentous algae cover as a nutrient response because the response can identify the potential for late-season deleterious hypoxia during the growing season when data are more easily collected.

Filamentous Algae Thresholds

The natural state of headwater streams in Utah is a state where stream autotrophs primarily consist of diatoms. DWQ included filamentous algae as a response based on the assumption that a regime shift to autotrophs dominated structurally in filamentous form is undesirable, and potentially detrimental to both recreational and aquatic life uses. The NNC threshold specifying that filamentous algae cover shall not exceed 1/3 of the stream bed was derived semi-quantitatively. The primary goal was to prevent filamentous algae from becoming the dominant form of benthic autotrophs, which is defined as >50%. This is because extensive filamentous algae cover is most strongly associated with adverse effects on aquatic life and also more likely to be associated with human-caused nutrient enrichment. Any value less than 50% could conceivably be considered protective against dominance of this growth form, but filamentous algae can accumulate—albeit to a lesser extent—in unenriched streams during periods of stable flow. To avoid making too many false impairment determinations, DWQ used 20% to demarcate an upper limit of naturally-occurring conditions. The NNC threshold (33%) was selected because it is intermediate to these lower and upper benchmarks.

Gross Primary Production

GPP is a direct measure of fundamentally important ecological alterations to nutrient enrichment and is the most proximate response to nutrient-related stress along most causal pathways leading to degradation of aquatic life uses. The GPP threshold in the NNC was the lowest GPP change point with respect to ambient nutrient concentrations that was statistically detectable. Linkages between stream metabolic rates and higher trophic levels are intrinsically indirect, but many nutrient-related adverse impacts to the composition of stream assemblages are initiated by increases in autotrophic (GPP) production. Compositional changes to stream autotrophs under nutrient enrichment result in a competitive advantage to taxa that are more competitive in enriched environments. At low to moderate level of nutrient enrichment, these changes are not always detrimental to higher trophic levels. However, as GPP rates increase, compositional changes can alter dominant growth forms, which cause changes in stream habitat (i.e., DO, pH, benthic habitat space) and other alterations to stream food webs. Many of the nutrient-related alterations to the composition of macroinvertebrates and diatom assemblages documented by DWQ (Chapter 7) are indirectly attributable to increases in GPP. At extremes, these changes affect sensitive fish via a reduction in active drift of macroinvertebrates and increase in adverse conditions (e.g., low DO, high pH). The NNC are intended to prevent degraded conditions by identifying GPP rates that are reflective of the early stages of inceased production that could degrade aquatic life if the continued increase was unabated.

Gross Primary Production Thresholds

The GPP threshold represents the lowest statistically valid change-pointsassociated with ambient nutrient concentrations. This response was interpreted as an inflection point that distinguishes background GPP rates from those altered by nutrient enrichment. The corresponding TN and TP concentrations are roughly equivalent of the 75th percentile of reference site distributions (TSD, Chapter 11), which provides some support for the interpretation with respect to the causal realtoinship with GPP.

Additive Protection through Combined Responses

Given the importance of auotrophic responses to enrichment on adverse effects to aquatic life that could ultimately occur, it was important that these responses were captured in the most comprehensive way possible. Both GPP and filamentous algae cover are included in the proposed NNC as autotrophic responses because both have strengths and weaknesses as indicators of excessive primary production. GPP measures a fundamental ecosystem process and provides a reach-scale measurement of primary production rates. Filamentous algae cover provides a measurement of the cumulative effects of nutrient loads throughout the growing season. The use of both indicators also provides flexibility with respect to the monitoring necessary to identify nutrient-related problems. Filamentous algae cover measurements can be made quickly and inexpensively, which means that they can potentially be collected during routine monthly monitoring to better understand within-growing season changes in algal accrual. Sondes that collect the time series of DO and water temperature necessary for metabolism calculations could hypothetically be deployed throughout the growing season. In reality, DWQ will need to rotate sondes among streams throughout Utah to meet ongoing monitoring demands. While efforts will be made to deploy sondes during conditions when excessive GPP and ER are most likely, estimates of when this may occur at a stream may not be accurate. Filamentous algae cover can be measured at least monthly, and having response data collected throughout the growing season may help DWQ identify nutrient problems that otherwise might be missed.

The selected GPP threshold was based on the S-R threshold that best distinguished between streams in good versus fair condition. It is possible that small excursions over this threshold may not be sufficient for a stream to be considered impaired; however, even under such circumstances these thresholds are appropriately protective given the ecological and economic importance of headwater streams. The proposed thresholds are also similar to those proposed by Young and colleagues (2008) who compiled metabolism data from streams throughout New Zealand and then used frequency distribution methods to define three condition classes. Streams in their study were considered to be in fair condition when GPP was between 4 and 8 g O_2/m^2/day. While the authors caution that their published condition classes should be confirmed locally, the fact that DWQ's proposed threshold of 6 g O_2/m^2/day is similar provides additional support that the conditions classes are broadly applicable.

Heterotrophic Responses

Direct measures of heterotrophic responses remain difficult to incorporate into routine monitoring and assessment programs. ER measures the net daily respiration in a stream and includes the respiration of both autotrophs and heterotrophs. Production rates of both autotrophs and heterotrophs can increase as ambient nutrient concentrations increase (Dodds 2007), making it difficult to differentiate which response causes an excursion of the ER component of the proposed NNC. For example, the S-R models for organic matter standing stock found that autochthonous biomass was significantly, albeit weakly, correlated with ER (Chapter 5). The relative contribution of autotrophs and heterotrophs to ER likely differs among streams. Scott and colleagues (2008) found that algal and bacterial production was closely coupled at streams with higher nutrient concentrations but became decoupled when nutrient concentrations were low. In contrast, Rier and Stephenson (2001) found that algal biomass was predictive of bacterial biomass in streams with higher nutrient concentrations but not among streams with lower nutrient concentration. Clearly, decoupling of algal and bacterial production can occur, but the conditions under which this occurs needs further investigation. Despite these limitations, ER remains the best measure of heterotrophic response to nutrient enrichment, and DWQ plans to incorporate it into ongoing monitoring and assessment programs.

Like GPP, the proposed ER threshold (5 g O_2/m^2/day) was based on the boundary for streams in fair versus poor condition developed from the S-R models. Also like GPP, this threshold is remarkably similar to the range of ER rates that Young and colleagues (2008) reported to be associated with their intermittent condition class: 5.5–10 g O_2/m^2/day. Another line of evidence that supports the proposed ER threshold is its relationship to Utah's numeric DO criteria. No violations of minimum daily DO criterion were observed at any sites in good condition, as measured with ER, but an increasing number of excursions of these criteria were observed at sites in fair condition, and many more excursions were observed as streams moved from fair to poor condition (Chapter 5). A significant increase in the number of excursions below the 30-day DO criterion were observed once ER reached rates associated with streams in fair or poor condition. A separate analysis found an ER of about 6 g O_2/m^2/day to be the point at which the proportion of DO observations that fell below the 30-day DO criterion transitions from consistently near zero to a much larger proportion (up to approximately 70% of all observations) as ER continues to increase (Figure 12.3).

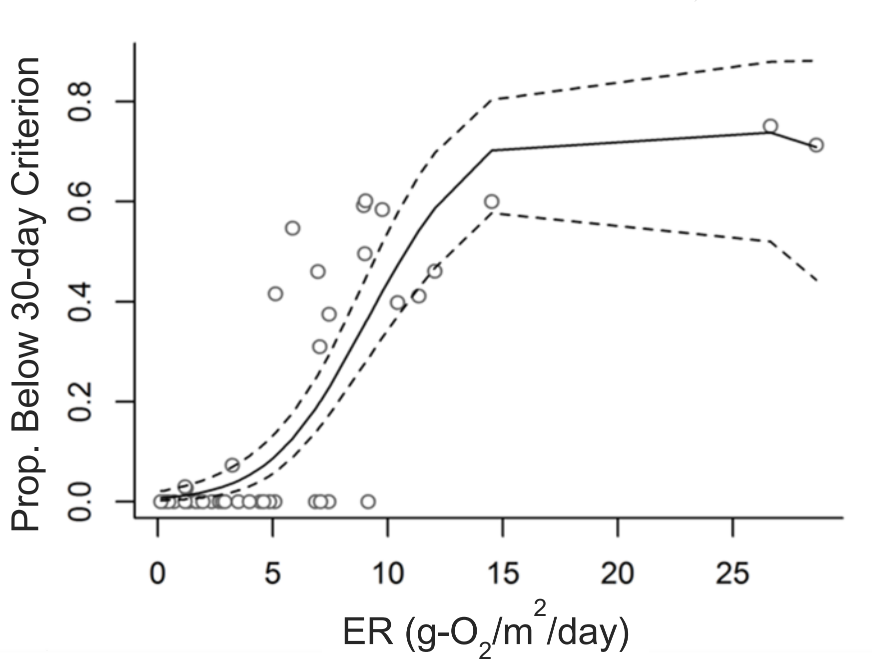


Figure 12.3. Relationship between ecosystem respiration and the proportion of site dissolved oxygen observations that fell below Utah's 30-day average dissolved oxygen criterion.

Relationships to Aquatic Life Uses

The primary objective in proposing these NNC is the protection of aquatic life uses. Maintenance of healthy ecological function is integral to meeting this objective, but in practice structural (compositional) indicators are used to assess biological use support. Biota at higher levels in stream food webs are indirectly affected by increased nutrient loads, which sometimes makes these indicators less sensitive to nutrient enrichment than more proximate responses (USEPA 2015). Nevertheless, the proposed NNC thresholds were compared with the results of investigations that demonstrated changes to the biological composition of stream biota resulting from nutrient enrichment. In particular, this TSD evaluated the likelihood of the lower nutrient thresholds missing a biologically degraded stream because of a lack of ecological response data; the lack of ecological data could occur because these data may not be routinely collected when ambient nutrient concentrations are below the lower nutrient thresholds.

The S-R models explored changes in the composition of macroinvertebrates, and to a lesser extent, diatoms (Chapter 7). Compositional changes of sensitive, tolerant, and all taxa with were identified with TITAN (Baker and King 2010). Diatom composition was found to be more sensitive to TP enrichment than macroinvertebrate composition, which is consistent with the results of others (Justus et al. 2000, Lavoie et al. 2008). Data were insufficient to evaluate diatoms for TN, but the TITAN model for all diatom taxa identified a threshold of greatest compositional changes at TP concentrations that ranged from 0.010 to 0.047 mg/L (5^{th}–95^{th} percentile of bootstrapped estimates). The TITAN model for all macroinvertebrates identified TP thresholds between 0.004 and 0.113 mg/L, and TN thresholds between 0.4–1.1 mg/L. The proposed TN and TP thresholds fall within these estimates, although they are in the upper end of the range for diatom taxa. More convincing support for the proposed NNC was provided by the second analysis that examined changes in macroinvertebrate taxa using a river invertebrate prediction and classification system (RIVPACS) model (Wright 1995) that DWQ currently uses to assess support of aquatic life uses. Streams were grouped into those meeting and not meeting aquatic life uses—as determined by independently developed models and impairment thresholds. Sites where TN was > 0.41 mg/L (95% confidence interval [CI] = 0.12 and 0.79) and TP was > 0.045 mg/L (95% CI= 0.023 and 0.066) were found to be the thresholds that best balanced false-positive and false-negative errors in impairment determinations. The lower threshold proposed in this TSD is nearly identical to the nutrient concentrations that best distinguished impairments when analyzed using an assessment method developed independently from the S-R analysis presented in this TSD.

Other studies have documented changes in the compositions of macroinvertebrates and fish in response to increasing nutrients; however, most do not derive or report threshold responses. Several exceptions are summarized here. Wang and colleagues (2007) evaluated changes in the composition of macroinvertebrate assemblages associated with streams with varying levels of ambient nutrients and found TP thresholds of 0.04–0.09 mg/L and TN thresholds of 0.6–1.6 mg/L, depending on the specific metric evaluated. The same authors evaluated effects on fish and found thresholds for salmonid metrics at ~0.6 mg/L TN and ~0.06 mg/L TP. Evans-White and colleagues (2009) found TN thresholds ranging from 0.93–1.14 mg/L and TP metrics ranging from 0.05–0.06 mg/L in their evaluation of different macroinvertebrate metrics. The lower thresholds for TN and TP proposed here fall below all these thresholds, which provides further support that the proposed NNC are likely to be protective of organisms at higher trophic levels.

Benchmarking against Criteria Proposed by Others

The proposed NNC thresholds were compared with criteria or other regulatory limits that have been proposed or adopted elsewhere. This analysis is potentially useful because, if the proposed NNC thresholds are similar to those derived by independent methods and analysis, then the proposed NNC may be reflective of conditions that are broadly protective of stream biota.

The lower thresholds proposed for TN and TP generally agree with TMDL endpoints established in Utah, which are often established using site-specific mechanistic models. The proposed NNC are also similar to those proposed or adopted for the protection of other streams in the intermountain west. Montana Department of Environmental Quality recently promulgated seasonal NNC for TN at 0.250–0.325 mg/L and—with one exception of an isolated volcanic range—from 0.025–0.030 mg/L for TP (Suplee and Watson 2013). In Colorado, stream nutrient criteria of 0.090 mg/L TP and 0.84 mg/L TN were recommended to protect cold water fish, although these have not yet been approved by USEPA. Other western states, like Arizona and California, currently only have TN or TP criteria for a limited number of streams, with values that are similar to those proposed by DWQ.

Conclusion

Elsewhere, NNC have been proposed that rely exclusively on water column nutrient concentrations. One concern with taking a similar approach for Utah streams is that an over-reliance on ambient nutrients could lead to under-protection of streams where production was the greatest because biological uptake rates increase with increasing nutrient concentration (Dodds et al. 2002). When water column nutrients are incorporated into suspended algae, the nutrients can be captured with samples that quantify total (both organic and inorganic) N and P. However, in small- to moderate-size streams most primary production is benthic, which TP or TN water column samples may miss. Nutrient spiraling also plays a role because organic nutrients in suspended algae cycle much more quickly to more biologically available inorganic forms than they do in benthic algae, especially if the benthic algae largely consist of taxa that accumulate over long periods of time (e.g., filamentous algae). These macroalgae can create zones of transient nutrient storage, further contributing to declines in observed ambient nutrient concentrations (Dodds 2003, Mulholland et al. 1994). The net result of this scenario could generate deceptively low nutrient concentrations at sites where the need to address nutrient enrichment is greatest.

Another advantage to combined NNC is that they should help minimize erroneous findings of impaired aquatic life uses. This is important because all the evaluated S-R models showed that covariates such as channel shading, stream slope, and stream size are almost as important as ambient nutrient concentrations in determining whether degradation was observed at a stream. This observation is consistent with every study that has evaluated factors influencing the ecological responses to nutrient enrichment and has been cited as one of the principal reasons why work on the trophic status of streams is much less developed than it is in lakes (Dodds 2007).

Chapter 13

AN INDEPENDENT EVALUATION OF THE PROPOSED NUMERIC NUTRIENT CRITERIA FOR HEADWATER STREAMS

Key Points

A study was conducted in 2015 to evaluate the proposed numeric nutrient criteria (NNC) for headwater streams.

Results are presented for 49 streams where data were obtained for both nutrients and at least one NNC response.

Average total P was higher than generally observed among headwater streams (0.58 ± 0.063 mg-P/L), whereas total N was approximately equivalent (0.34 ± 0.20 mg-N/L).

Approximately 14% of the sites exceeded the upper NNC threshold, 6 for P and 1 for N.

Stream metabolism data suggest that these stream were more sensitive to increases in ecosystem respiration (ER) (average: 0.04 – 9.83 g O2/m2/day) than gross primary production (0.01 – 3.37 g O2/m2/day), although one stream exceeded the NNC threshold for ER.

Filamentous algae proved to be a more sensitive response with 8 streams exceeding the NNC threshold on one or more site visits.

The use of both nutrients and responses at streams with low to moderate levels of human-cause d nutrient enrichment was useful in separating stream where potentially deleterious responses occurred from those stream where local conditions precluded negative affects to aquatic life uses.

Errata

Metabolism data for two study sites, East Fork Virgin River at US89 (Site 4951550) and Huntington Creek (Site 4930585) were flagged as having suspect stream metabolism results and should have been removed from assessment tables. Field observations at both sites suggest that metabolic rates were influenced by turbidity leading to relative inaccurate estimates of GPP and ER. Metabolic data for both sites should be considered non-detects and site-specific data in summary tables have been flagged accordingly. The authors regret missing the removal of these data points through internal quality control processes.In both cases, the ultimate interpretation of the data with respect to the NNC remain unchanged because GPP and ER are well below proposed NNC thresholds.

Introduction

In order to derive headwater numeric nutrient criteria (NNC), a diverse set of ecological responses were evaluated over a broad range of nutrient concentrations to better understand and quantify relationships between nutrients and alterations to stream biota that, left unchecked, could ultimately lead to the degradation of aquatic life uses. The stressor-response (S-R) models that were generated through these investigations helped provide the evidence needed to determine NNC to protect aquatic life (Chapter 12) and recreational (Chapter 8) uses. When considered together, these studies were sufficiently robust to develop defensible headwater NNC, yet these studies also highlighted a number of shortcomings that needed to be addressed to ensure that these criteria could be easily implemented after their adoption into Utah's water quality standards. For instance, the dataset used to derive the proposed NNC did not include a large number of streams in Utah's headwaters. Nutrient concentrations in the headwaters are lower and have a smaller range of variation than streams in the dataset used to derive the S-R relationships. In addition, ongoing discussions with cooperating agencies revealed a need to develop and refine the field methods used by Division of Water Quality (DWQ). To address knowledge gaps, in 2015 DWQ instigated a cooperative project with the Utah Division of Agriculture and Food and the U.S. Forest Service (USFS) to collect nutrients and ecological response data from as many headwater streams as possible. The results of this investigation are described in this chapter of the technical support document (TSD).

One of the most common ways that excessive production is manifest in Utah's headwater streams is a shift from an algal assemblage dominated by diatoms to one dominated by less desirable filamentous algae, particularly *Cladophora* (Dodds and Gudder 1992, Chetelat et al. 1999). However, filamentous algae cover response was not evaluated with the S-R models that were initially developed to support NNC development. Instead, the S-R investigations used whole stream metabolism and more conventional measures of algal biomass—benthic chlorophyll-*a* (chl-*a)* and ash free dry mass (AFDM)—to quantify increases in primary production caused by nutrient enrichment (Hauer and Lamberti 2011). While these measures of production remain robust, they each have shortcomings that were complicating efforts to scale up data collection efforts needed to sufficiently support statewide NNC criteria. Gross primary production (GPP) is an easy and inexpensive measure of primary production but has its own limitations. Prices for the sondes needed to collect the data required for metabolism models continue to decrease, but DWQ will never have enough equipment to deploy sondes at all headwater streams. In addition, there are circumstances in which metabolism models cannot accurately calculate GPP and ecosystem respiration (ER), because diel changes in dissolved oxygen (DO) are insufficient for making reliable reaeration estimates. Such curcumstances are not problematic with respect to the protection of uses because GPP and ER would also be far below NNC thresholds. This is akin to a non-detect analytical result in water chemistry data. Nevertheless, increases to primary production are among the most important responses to nutrient enrichment, so the inclusion of additional responses in the NNC was desireable. The search for alternative measures of primary production identified a handful of regulatory agencies that estimate filamentous algae cover using methods that could be routinely incorporated into statewide monitoring efforts (Stevenson et al. 2012, West Virginia Department of Environmental Protection). The addition of this measure to the proposed headwater NNC was appealing because it could be used to identify streams where more resource-intensive follow-up monitoring was warranted. However, DWQ needed to adapt and test methods to collect this response, which became an objective of this investigation.

The overall objectives for this investigation included the following: (1) obtain ecological response data from as many headwater sites as possible where previous nitrogen (N) or phosphorous (P) concentrations exceeded the proposed lower or upper NNC thresholds, (2) estimate the implementation costs of the proposed rules through better estimates of the frequency of impairments related to excess nutrients in headwater streams, (3) develop and refine field and laboratory methods for measures that had not previously been part of DWQ's routine monitoring and assessment programs, and (4) confirm that the proposed combined criteria are appropriately protective of the beneficial uses of headwater streams.

Methods

Study Locations

Site selection efforts began by compiling and summarizing existing nutrient data for all samples collected from Utah's headwater streams over the 9 years prior to the study's initiation (approximately 450 unique locations, 2004–2014). Summertime total N (TN) and total P (TP) concentrations were averaged, and candidate sample locations were identified as any stream where TN or TP exceeded the lower or upper thresholds of the proposed headwater NNC. These candidate sites were mapped, and a final list of 49 sites was generated with the dual goal of including as many headwater streams with high concentrations of TN and TP as possible, while also ensuring that sample locations were spatially dispersed (Figure 13.1).

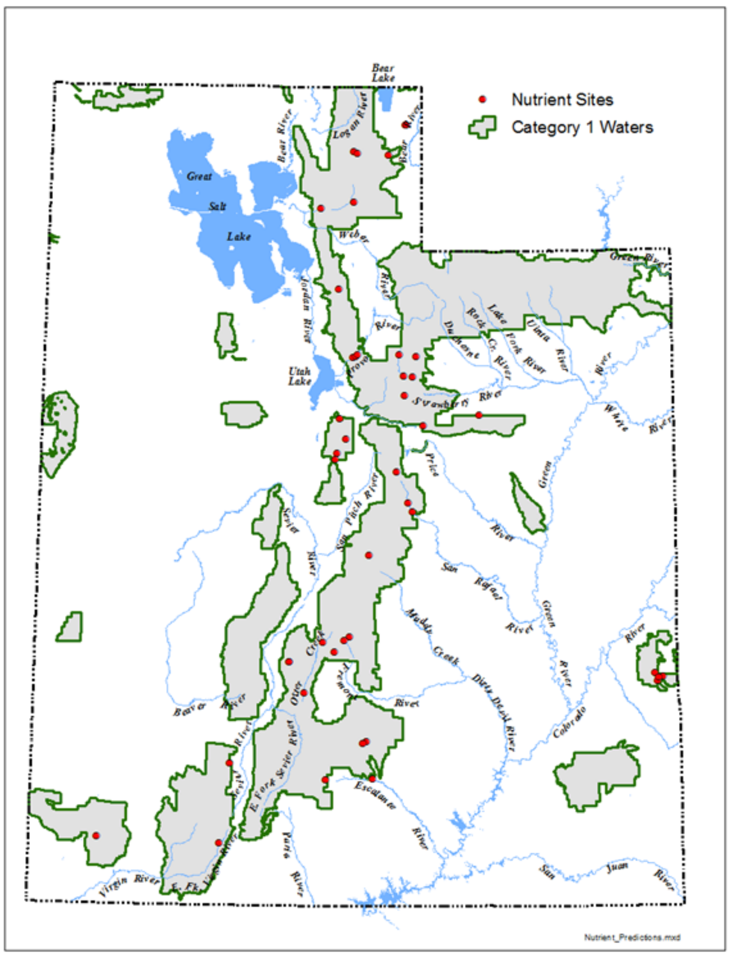


Figure 13.1. Map of 2015 study locations.

Data Collection

Data were collected at each of the study locations approximately monthly from June through September in 2015. During the first visit, a stream reach approximately 100 m long was established and georeferenced; three transects were established within the reach at approximately the top, bottom, and midpoint of its length. Transect were demarcated by hammering rebar into the stream bank and georeferenced so that field personnel could relocate them on subsequent site visits.

Water Chemistry

Grab samples for water chemistry were collected immediately upon arrival at each site. Grab samples for both filtered and unfiltered nutrients were collected following DWQ standard operating procedures.[1] Samples for dissolved analytes were field filtered through 0.45 µm membrane filters (Millipore Corporation). Nutrient chemical analyses were conducted by the Utah Health Laboratory for total N, total DO, total P and total dissolved P using persulfate oxidation followed by standard colorimetric analysis (Valderrama 1981). The lab also processed these samples with standard colorimetric analysis to obtain nitrate + nitrite (USEPA method 353.4), ammonium (USEPA method 349), and P (USEPA method 365.5).

Metabolism

Data Collection

The principal data necessary for whole stream metabolism models are a high frequency time series of DO and temperature. Largely due to equipment limitations, these data were obtained for only 7–10 days at each study site. At each study location a MiniDO$_2$T (PMI instruments®) was deployed that recorded temperature and DO in 5–15-minute intervals over the period of its deployment. Independent measures from a recently calibrated sonde were also collected at the end of each deployment to evaluate whether sensor fowling resulted in parameter drift.

In addition to DO and temperature, metabolism model calibration also requires the collection of stream discharge, slope, water depth, and several measures of climate conditions. Standard DWQ procedures were followed to obtain discharge measurements (width, depth, and current velocity) using a FlowTracker Handheld Acoustic Doppler Velocity meter (SonTek) when each sonde was deployed. Discharge was calculated using the velocity-area method (Levesque and Oberg 2012). Slope was measured using a distance finder and transit survey level over several segments of the stream reach. Weather parameters needed for model calibration—air temperature, barometric pressure, and light radiation—were obtained from the nearest weather station in the Mesowest weather data repository that was at an elevation closest to the stream where the sonde was deployed.

Model Construction

The R (R Core Team 2012) library streamMetabolizer (Appling et al. 2017) was used to create metabolism models using an open water method with reaeration (K) as a free parameter. Daily GPP and ER were derived based on the relations defined by the
following equation from Van de Bogert and colleagues (2007):

[1] https://deq.utah.gov/legacy/monitoring/water-quality/quality-assurance-quality-control.htm

$$O_t = O_{t-1} + (\frac{GPP \cdot \Delta t}{z} \times \frac{Light_t}{\sum Light} + \frac{ER \cdot \Delta t}{z} + K(O_{sat} - O_{t-1}) \cdot \Delta t)$$

Where,

ER = Ecosystem respiration (loss of g O_2/m^2/day)
GPP = Gross primary production (g O_2/m^2/day)
K = Reaeration coefficient (day^{-1})
Light = Solar radiation or photosynthetic active radiation (PAR)
O = Dissolved oxygen (mg/L)
O_{sat} = Oxygen saturation (mg/L)
t = Time (fraction of day)
z = Mean stream depth (m)

Additional details about the application of this equation to whole stream metabolism models are provided Chapter 4 of this TSD.

While the mechanics of metabolism models can be complicated, they all essentially work from an observation first proposed by Odum almost 70 years ago: DO observed in the water column at night is a function of loss of oxygen from ER and atmospheric exchange (reaeration), whereas DO in the daytime also includes oxygen inputs from photosynthesis (Odum 1956). The difficulty lies in estimating reaeration rates (K) and appropriately spatially and temporally scaling the models to generate reach-scale estimates of GPP and ER. To parameterize the models, this study used a time series of DO and water temperature recordings from sondes deployed for 2–7 days at each location, field measurements of depth, and light measurements and air temperature measurements obtained from nearby weather stations (Mesowest database). Light intensity at the weather stations was measured as short-wave radiation, so this was converted to PAR using the function "calc_light" in the streamMetabolizer R package. DO saturation was calculated for each time step using water temperature obtained from the sondes and air pressure estimates derived from site elevation and air temperature. Metabolism models were generated that were the best possible fit to diurnal DO measurements using maximum likelihood estimation curve fitting. The resulting measures of GPP and ER are expressed as g O_2/m^2/day, positive for GPP, and negative for ER. Although both are reported as positive numbers, it should be noted that GPP measures the production of oxygen, and ER measures the loss of oxygen from the stream.

Algal Abundance

Data Collection

A visual measure of filamentous algae was obtained across each of three transects at each study site during each of the four collection events using a line-intercept method. First, each transect was subdivided into 3–10 habitat units (HUs) that were relatively uniform in physical characteristics (e.g., flow, depth, substrate size). The width of the HU was measured, and the width of filamentous algae cover was measured at the center of the HU (from bank-to-bank). The presence or absence of filamentous algae was measured by holding a PVC rod marked in ten 10-cm intervals perpendicular to the flow; the presence or absence of algae was recorded for each 10-cm mark, and for approximately 5 cm to either side of each 10 cm mark. These measurements were taken at each HU looking both upstream and downstream, resulting in as many as 20 recordings of filamentous algae cover for each HU. When the stream bed was not clearly visible, this was noted, and the data were not included in either the numerator or the denominator for purposes of percent filamentous algae cover calculations. Samples were collected to quantify benthic biomass; these were preserved by freezing the samples until they could be processed. Unfortunately, the freezer malfunctioned before laboratory processing, so these data are unavailable.

Algae Cover Calculations

A reach-wide measure of filamentous algae cover was calculated as an average of the measures for each of the three transects, at each stream, for each collection event. Because the number and characteristics of HUs differed within and among streams, each HU measurement was expressed relative to the width of the HU relative to the total wetted width of the transect. For example, if the width of an HU was 20% of the transect width, then the percent filamentous algae cover measured for that HU would similarly contribute 20% to the transect average. The algae cover measurements for the three transects were averaged to generate a reach-scale filamentous algae cover measurement. The maximum filamentous algae cover recorded was extracted from among the four site visits since the proposed criterion is written as a seasonal not-to-be-exceeded criterion.

Results

Sampling Logistics

The collection season was started with the aim of collecting data at approximately 70 sites. However, due to logistical constraints associated with sampling these remote locations, the sampling schedule was modified throughout the field collection season. As a result, data presented in this chapter are constrained to 49 sample sites from which sufficient data were collected to calculate growing-season nutrient concentrations and at least one response (i.e., metabolism or filamentous algae cover) (Figure 13.1).

Despite the fact that study sites were constrained to headwater streams, they exhibited a fairly wide range of physical characteristics. While sites were all at relatively high elevations (average of ~2,200 m) relative to all streams in Utah, they spanned a fairly wide range of elevation, from 1,580 to 3,040 m. Canopy cover (shading of the stream channel by riparian vegetation) averaged about 50% but ranged from 3% to 87%. Sites were generally of high gradient, averaging about 5% slope, with a range from < 0.5% to just over 20% among all streams. Given these relatively high gradients, it is not surprising that about 80% of sites had relatively coarse substrates (> 50% of measured cobbles were 64–250 mm or larger).

Nutrient Concentrations

The objective of selecting headwater sites with high nutrient concentrations was more successful for TP than TN (Figure 13.2). Among all sites, growing-season average TP was 0.050 ± 0.063 mg-P/L, ranging from a low of 0.005 mg-P/L to a maximum of 0.329 mg-P/L. This means that TP at these study sites was, on average, about 66% higher than the average observed among all headwater streams (Chapter 11). In contrast, growing-season average TN was 0.34 ± 0.20, and ranged from 0.09 to 0.20 mg-N/L, which was almost identical to expectations based on the broader analysis of headwater nutrient concentrations (0.33 mg/L, N = 448).

The proposed criteria specify two nutrient thresholds: an upper threshold where nutrient concentrations alone could be used to document an impairment of aquatic life uses (TP > 0.080 or TN > 0.80 mg/L) and a lower concentration where ecological confirmation is required prior to concluding an impairment of aquatic life uses (TP = 0.035-0.080, TN = 0.40-0.80 mg/L; Chapter 10). Based on these criteria, 2% (1 of 49 sites) of study locations would be considered impaired due to exceeding the upper threshold for TN, and 14% of sites (6 sites) would be considered impaired for TP. No study location exceeded the upper threshold for both TP and TN. In comparison with the proposed lower thresholds, an additional 10 sites would require additional evaluation on the basis of TN, and an additional 17 sites would

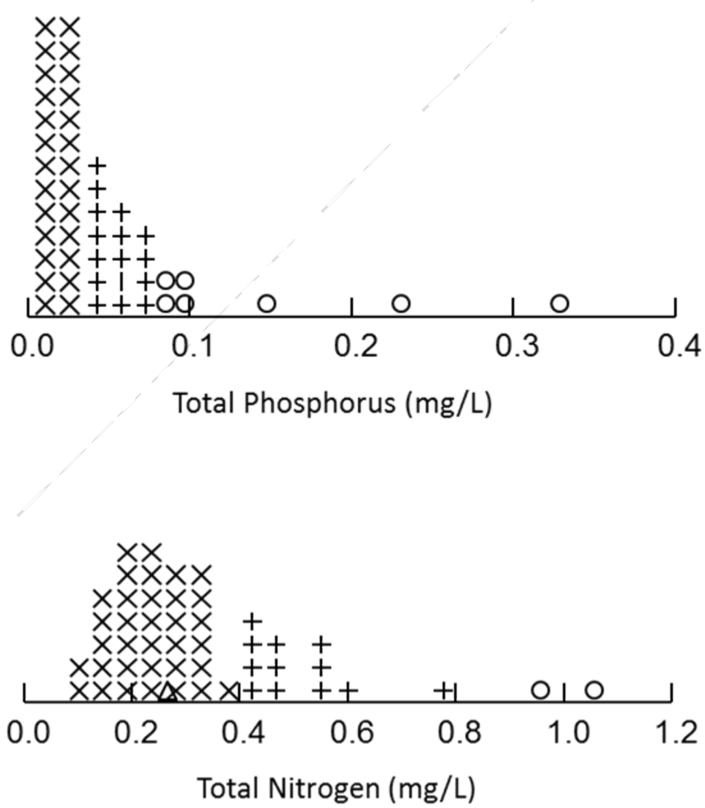


Figure 13.2. Average phosphorus and nitrogen concentrations observed among study sites, coded with respect to proposed numeric nutrient criteria thresholds: X = below the lower threshold, + = between thresholds, and o = above upper.

require further evaluation on the basis of TP.

Metabolism

Several circumstances limited the construction of metabolism models at the 49 sites. At some sites, diel variation in O_2 was so low that accurate reaeration estimates could not be generated. In addition, data from some sondes were either not recorded or were incorrectly extracted from the instruments. As a result, metabolism models are only available for 31 of the sites in this investigation.

In general, neither GPP nor ER was high in comparison with the broader investigation used to generate S-R relationships (Chapter 5). GPP ranged from 0.01 to 3.37 g O_2/m^2/day, and ER ranged from 0.04 to 9.83 g O_2/m^2/day. With respect to the proposed impairment thresholds, no site exceeded the threshold for GPP (> 6 g O_2/m^2/day); one site exceeded the threshold for ER (> 5 g O_2/m^2/day), and two additional sites were close to the ER threshold (Table 13.1).

Table 13.1. Average nutrient (total nitrogen [TN], total phosphorus [TP]) concentrations and associated ecological response measurements at all study sites. For sites where the proposed headwater numeric nutrient criteria would indicate impaired aquatic life uses, the reason for the impairment is provided.

MLID	Site Description	Reason for Imp.	Responses			Nutrients	
			GPP	ER	Fil. Algae	TP	TN
4930667	Crandall Creek above confluence with Huntington Creek	--	0.37	0.61	16	0.006	0.49+
5931820	Indian Creek above Ferron Creek Reservoir	--	0.69	0.47	0	0.007	0.09
4995840	Holman Creek above confluence with Nebo Creek	--	0.40	2.22	10	0.008	0.24
4905500	Blacksmith Fork above Hardware Ranch	--	1.38	2.85	3	0.008	0.27
4954650	E Fork of the Virgin River above confluence with Stout Creek	--	--	--	20	0.009	0.20
4958883	Beaver Creek above Chicken Creek Ditch	--	--	--	0	0.009	0.21
4958877	Deer Creek Spring below Confluence	--	--	--	4	0.009	0.26
4905480	Blacksmith Fork River below Hardware Ranch	--	3.00	3.21	0	0.009	0.33
4954210	Calf Creek above confluence with Escalante River	--	0.46	0.63	15	0.010	0.13
4996850	N Fork Provo River above confluence with Provo River at Wildwood	–	0.44	4.80	1	0.011	0.33
4996830	Lower S Fork Provo River at Gaging Station	--	0.11	0.36	0	0.012	0.33
4924700	S Fork Ogden River below Causey Dam Spillway USFS	Fil. Algae, TN	1.51	1.81	44	0.012	0.44+
4905482	Big Spring	Fil. Algae, TN	--	--	89	0.014	0.58+
4924650	N Fork Ogden River at USGS Gage below Bridge on U162 USFS	--	1.38	2.85	1	0.016	0.26
4936510	Trout Creek above Strawberry Reservoir	TN	0.78	0.04	0	0.018	0.96++
4908193	Big Creek above exclosure and private land	--	--	--	23	0.019	0.33
4936610	Indian Creek above Strawberry Reservoir	--	0.69	0.47	1	0.020	0.14

MLID	Site Description	Reason for Imp.	Responses			Nutrients	
			GPP	ER	Fil. Algae	TP	TN
4956470	LaSal Creek near Headwater	--	--	--	0	0.020	0.19
4905478	Big Creek at USFS Boundary	--	--	--	0	0.021	0.38
4948881	Otter Creek ~1.5 miles above Rock Canyon (UT09St-761)	ER, Fil. Algae	1.59	9.83	95	0.022	0.27
5960320	Vernon Creek below private land	--	--	--	28	0.024	0.26
4953942	Birch Creek below confluence with Corn Creek	--	--	--	17	0.025	0.15
4995380	Salt Creek above confluence with Red Creek	--	0.74	1.90	32	0.026	0.24
4908148	Sage Creek above confluence with North Fork of Sage Creek	--	--	--	0	0.026	0.60+
4949591	Antimony Creek	--	--	--	6	0.028	0.28
4931250	Huntington Creek above Electric Lake	--	1.15	2.46	0	0.034	0.54+
4953920	West Fork Boulder Creek above confluence with East Fork	--	--	--	15	0.038+	0.13
4958890	LaSal Creek above Road 073 crossing		0.64	4.91	30	0.039+	0.22
4945795	Otter Creek at USFS boundary	--	0.21	1.02	0	0.041+	0.31
4936620	Clyde Creek	--	0.19	1.00	15	0.043+	0.41+
4954960	Twin Creek above Fish Lake	--	--	--	22	0.043+	0.22
4992170	Mountain Dell Creek at U65 crossing below Little Dell Reservoir	--	0.64	0.46	0	0.045+	0.22
4996810	Provo River at Olmstead Diversion	Fil. Algae, TP, TN	0.10	0.20	87	0.048+	0.45+
5956150	Seven Mile Creek above Johnson Reservoir	--	0.06	0.20	7	0.049+	0.27
4951550***	E Fork Virgin River At US89 S Mt Carmel Junction	TN	3.37	0.15	14	0.049+	1.06++
4995360	Salt Creek at USFS boundary	--	--	--	32	0.053+	0.41+
5936480	Currant Creek above Pass Creek	--	0.48	1.59	30	0.060+	0.27
4953970	Bear Creek below Haws Pasture	Fil. Algae, TP	0.14	0.63	40	0.060+	0.34
4995510	Peteetneet Creek above Maple Dell campground	Fil. Algae, TP	--	--	68	0.063+	0.20
5956000	Um Creek At USFS Road 015	Fil. Algae, TP	0.05	0.11	52	0.067+	0.24
4936685	Strawberry Reservoir below E Daniels Allotment	--	1.23	2.18	7	0.068+	0.26
5945030	Manning Creek below Manning Meadows Reservoir	--	--	--	30	0.079+	0.32
4908150	Sage Creek below confluence with North Fork of Sage Creek	Fil. Algae, TP	--	--	33	0.081+	0.46+
4949730	Threemile Creek 0.75 miles W of USFS Boundary	TP	0.15	1.06	0	0.087++	0.18
4932880	Right Fork White River at USFS boundary	TP	1.13	1.40	22	0.087++	0.27
5950595	Santa Clara River below Pine Valley	TP	0.14	0.33	18	0.098++	0.55+
4930585***	Huntington Creek ~0.5 miles below Trail Canyon	TP	3.08	1.53	3	0.148++	0.31
5945070	Timber Creek above Manning Meadows Reservoir	TP	--	--	3	0.231++	0.22

MLID	Site Description	Reason for Imp.	Responses			Nutrients	
			GPP	ER	Fil. Algae	TP	TN
4936000	L Fork Indian Canyon above USFS boundary	TP	0.01	0.58	9	0.329++	0.78

Notes: Gross primary production (GPP) and ecosystem respiration (ER) are expressed as g O_2/m^2/day, filamentous algae (Fil. Algae) is expressed as % of stream bed, and nutrients (total nitrogen [TN], total phosphorus [TP]) are expressed a mg/L. Responses that violated proposed numeric nutrient criteria (NNC) thresholds are in red text. For TP and TN, + = nutrient concentrations above the lower NNC threshold, and ++ = nutrient concentrations above the upper NNC threshold.Two sites (MLID = ***) were flagged as having suspect GPP and ER rates in peer review and should have been considered non-detects with respect to both responses.

Filamentous Algae

On average, the maximum filamentous algal cover observed was low but highly variable among study locations (average of 19 ± 24; Figure 13.3). At approximately 50% of the study sites (24 of 49) filamentous algae never exceeded 10% during the growing season. Among sites that had greater algal accrual, 8 sites (16%) had sufficient algal cover to exceed the proposed NNC threshold of 33% (Table 13.1). There was no statistically significant linear relationship between TN or TP and filamentous algae cover. Among sites where the growing-season maximum cover exceeded the proposed 33% NNC threshold, only one site did not exceed the lower nutrient threshold for both TP and TN. The relative importance to TN and TP appears to be roughly equal among sites that exceeded the filamentous algae threshold: 5 of 8 sites exceeded the threshold for TP, 3 of 8 sites exceeded the threshold for TN, and 2 sites exceeded for both TN and TP lower thresholds). Only one site where algal accrual exceeded the threshold failed to also exceed the lower threshold for both TN and TP.

Final Interpretation of Both Nutrients and Responses

Overall, if the proposed NNC were applied to these sites, 35% (17 of 49) would be considered to

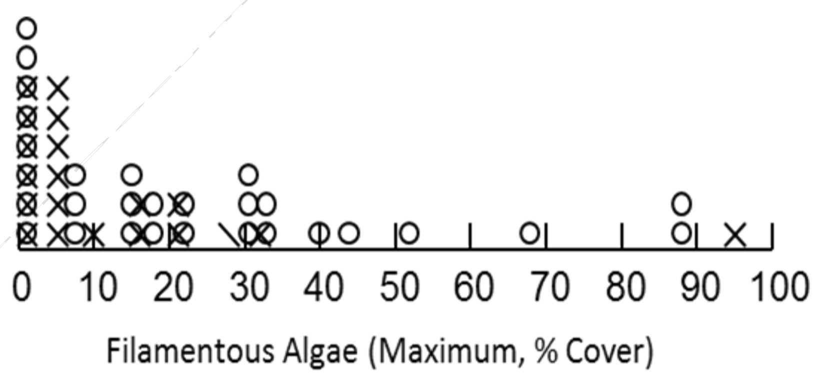


Figure 13.3. Maximum filamentous algae cover observed at each study location with streams coded with respect to average ambient nutrient concentration: 0 = sites where total nitrogen or total phosphorus exceeded the proposed lower numeric nutrient criteria threshold and X = sites where both total nitrogen and total phosphorus were below the lower threshold.

be impaired (failing to meet aquatic life uses). Among these impaired sites, 31% (7 of 17) would receive an impairment on the basis of nutrient concentrations alone (excursion of the upper threshold)—6 for TP and 1 for TN (Table 13.1). The 10 other impairment determinations (59% of all impaired sites) would result from a circumstance where TN or TP exceeded the lower threshold and a deleterious biological response was observed.

This evaluation of these hypothetical impairment determinations using both nutrients and ecological responses highlights the value of the combined criteria approach. Among the 16 sites that exceeded the lower—but not upper—TP threshold, 63% (10 of 16) would still be considered to be supporting their aquatic life uses. Similarly, among the 11 sites that required further evaluation because TN exceeded the lower threshold, 64% (7 of 11) did not exhibit any ecological response to elevated nutrient concentrations. Only one site had ecological responses that exceeded the proposed ecological thresholds without exceeding the lower TP or TN threshold, and this site also happened to have the greatest magnitude of excursion of both ER (9.83 g O_2/m^2/day) and filamentous algae cover (95%).

Discussion

Overall, this investigation supports the proposed NNC for headwater streams and the integration of ecological responses in determining use support. Few circumstances were identified where ecological responses exceeded the proposed NNC thresholds without a corresponding excursion of the lower TN or TP thresholds. This suggests that the lower values are appropriately protective. The results of this investigation demonstrate the value of incorporating ecological responses in the NNC; relying solely on the lower nutrient thresholds would have led to erroneous conclusions that some streams were not meeting their uses. The results also provide support for the use of filamentous algae cover as an ecological response in the NNC and, although whole stream metabolism responses were not as sensitive, there was not a compelling reason to eliminate this response from the proposed NNC. More broadly, this study confirmed some of the underlying assumptions and considerations used to create the proposed NNC, although it conflicted with other assumptions. The study results and their relation to the proposed headwater NNC are discussed below.

Insight into the Total Nitrogen and Total Phosphorus Numeric Nutrient Criteria Thresholds

For NNC to be protective of uses, they need to minimize the number of impairments that would be missed by the criteria. One way these Type II errors could potentially occur would be a circumstance where both TN and TP fall below the lower TN and TP thresholds, but other evidence demonstrates that nutrient-related degradation of the stream biota has occurred. Technically, such an error cannot happen with the proposed NNC because any excursion of a response threshold is considered a violation of the standard, although such scenarios could occur if low nutrient concentrations were found initially and response data were not collected subsequently. This investigation found one circumstance where an ecological response exceeded the proposed threshold and TN and TP did not. This site, *Otter Creek above*

Rock Canyon, was the most extreme example observed, with low nutrients (TN = 0.27 mg/L, TP = 0.022 mg/L) and the highest filamentous algae cover (95%) and ER values (9.83 mg DO/m^2/day). This is likely a circumstance where ambient nutrient concentrations were low due to high biological uptake rates and resulting nutrient retention. Overall, it appears that the proposed NNC are unlikely to miss problems associated with nutrient enrichment in headwater streams. Moreover, filamentous algae problems were so extensive at Otter Creek that follow-up investigations may have been conducted, irrespective of background nutrient conditions.

It is also important to minimize the number of circumstances where the NNC would identify an impairment that cannot be verified using ecological response data. For streams with nutrient concentrations between the upper and lower thresholds, the proposed NNC attempt to minimize these Type I assessment errors by requiring an evaluation of both nutrients and ecological responses. In all, 22sites were identified where the growing-season average of TN or TP fell between the upper and lower thresholds, and fewer than one-third (7) of these streams exhibited any signs of biological degradation. This observation highlights the value of combined criteria. Setting nutrient thresholds at concentrations that are low enough to identify impairments is important, and the proposed NNC attempt to do that with the lower thresholds. However, not including responses, and relying exclusively on the proposed lower NNC thresholds would have led to an erroneous impairment conclusion over 50% of the time.

Headwater streams with the highest known nutrient concentrations were intentionally selected this investigation. Despite this study design, streams that exceeded the upper nutrient thresholds were relatively rare. Only about 4% (2 of 49) of the study sites exceeded the upper thresholds for TN, while about 12% (6 of 49) of the study sites exceeded the upper threshold for TP. Since the upper TP threshold roughly corresponds to approximately the 95th percentile of TP observed among all headwater streams (Table 11.3), the observed excursion of this threshold is over twice what is expected (5%).

Only 1 of the 8 sites that exceeded an upper nutrient threshold also exceeded a proposed NNC threshold for any ecological response. Six of the sites that exceeded the upper threshold but did not exceed an ecological response level were related to high TP, while two were for high TN. In several cases, these sites occurred in streams where field crews noted evidence of late-summer, flash floods caused by monsoonal rain events. Such flooding events could both increase nutrient concentration and have minimized the magnitude of responses observed at these streams. Large flash floods can contribute a considerable amount of sediment and P from the landscape to streams (Verheyen et al. 2015). While much of the P may be particulate, it can be chemically or enzymatically hydrolyzed to more biologically available forms once it enters streams (Correll 1998), which would increase water column TP concentrations. Large flooding events can also scour algal biomass (Grimm and Fisher 1989), which would decrease filamentous algae cover and potentially reduce GPP and ER rates as well. Depending on the frequency and intensity of these events, the algal recovery periods, which range from a couple of weeks to several months (Stevenson 1990), may not have been sufficient to cause an excursion of response thresholds. Of course, it is also possible that other site conditions precluded the degradation of ecological responses at these sites. Many covariates can alter the extent of biological responses to nutrient enrichment. For example, unlike most of the streams in this investigation, many of these streams had beds that principally consisted of fine substrates. In general, algal biomass is generally greatest in streams with larger substrates (Biggs 1995). Light limitation may have been another factor limiting production at these sites, because many of the study sites had fairly high canopy cover, which in small streams can prevent the accumulation of filamentous algae (Lowe et al. 1986).

Consideration of Proposed Ecological Responses

Metabolism

Measures of biological integrity include consideration of both ecosystem structure and function (Bunn and Davies 2000, Norris and Hawkins 2000). However, in practice most indicators that are used in water quality management have been based on alterations to ecosystem structure (e.g., changes in the composition of fish or macroinvertebrate assemblages) resulting from human-caused stress to aquatic ecosystems. Ecosystem functions that describe the underlying processes of energy and matter are considered to be important but have generally been assumed to be protected as long as elements of structural integrity of aquatic ecosystems remain intact; this assumption has not been widely tested (Bunn et al. 1999, 2001). Several investigators have advocated for the incorporation of whole stream metabolic rates into biological assessment programs (Grace and Imberger 2006, Young et al. 2008). GPP and ER have been demonstrated to be indicators of ecological attributes that are reflective of stream health such as nutrient processing (Hall and Tank 2003) and ecosystem structure (Sabater et al. 2002). In addition, regional comparisons of metabolism have been shown to be responsive to landscape-level disturbances that cause increases in stream nutrient concentrations such as land use practices (Young and Huryn 1999) and riparian disturbance (McTammany et al. 2007). When GPP and ER rates were compared across the range of nutrients that occur throughout Utah, GPP and ER were found to be relatively robust indicators of enrichment, particularly when channel shading and slope were accounted for (Chapter 5). However, in the lower ranges of nutrient concentrations in headwater streams, these responses were not particularly sensitive measures of enrichment. This follow-up investigation did not document a single stream with elevated GPP, and one stream with elevated ER was identified.

In many cases, preliminary stream metabolism models could not accurately identify circumstances where metabolism models could accurately measure GPP and ER. Such circumstances are generally a result of an inability to estimate atmospheric reaeration, due to relatively small diel variations in DO. This is caused by circumstances that make streams naturally tolerant of nutrient enrichment, such as high turbidity, high canopy cover, high slope, low water temperature, and unconsolidated stream substrate. Under these circumstances, both ER and GPP are low, and although unquantified, can be concluded to be below the NNC thresholds. These circumstances are analogous to non-detects below the criterion in chemical analyses when the concentration of a pollutant could not be measured because concentrations are below the analytical detection limit.The fact that elevated levels of ER, but not GPP, were observed is supported by ecological theory and other observations. Almost 40 years ago, Vanote and colleagues (1980) proposed the river continuum concept (RCC) as a conceptual model of upstream to downstream changes to stream ecosystems. Among other things, this model predicts that upstream locations should be net heterotrophic (ER > GPP); the streams should then transition to a net autotropic condition (GPP > ER) further downstream. This shift is caused by a reduction in the relative importance of allochthonous organic matter inputs and an increase in the relative importance of autochthonous energy input as the canopy opens and stream primary production shifts from the benthos to the water column. Experimental support for this conceptual model includes a nutrient enrichment study conducted in headwater streams that found little influence of moderate nutrient enrichment to algal composition and biomass (Greenwood and Rosemond 2005).

Of critical importance to evaluating the proposed NNC is whether GPP and ER remain valid indicators of anthropogenic enrichment in Utah's headwater streams. As previously mentioned, GPP and ER are measures of fundamental ecosystem processes and integral aspects of biological integrity—the fundamental objective of the proposed headwater criteria. While a large number of GPP or ER threshold excursions were not observed, this does not discount the ecological importance of these processes. The proposed thresholds could be lowered to increase the sensitivity of ER, but it remains unclear that lower rates are necessarily reflective of biological impairments. In contrast, the proposed NNC thresholds were demonstrated to be protective from excursions to Utah's DO criteria (Chapter 4). It is possible to improve the sensitivity of ER responses by extending the index period to later in the year when ER may be higher as algae that accumulated during the growing season senesce and decompose. While worthy of exploration, it is important to note that many headwater streams may have physical characteristics that make them less susceptible to the deleterious biological responses caused by nutrient enrichment.

Filamentous Algae

The results of this investigation suggest that filamentous algae cover can be an effective response metric for headwater streams. While not all streams where N or P were above background concentrations (i.e., higher than the proposed lower thresholds) had elevated levels of filamentous algae cover, nearly all streams with elevated algae cover had elevated nutrient concentrations. The fact that not all streams with elevated nutrients had high filamentous algae supports the combined use of both nutrients and responses in the proposed NNC. This approach has the added benefit that measures of filamentous algae cover are readily incorporated into routine monitoring programs.

Combined Consideration of Nutrients and Responses

This study confirms the use of combined nutrient criteria for Utah's headwater streams. The lower nutrient thresholds proposed by DWQ are conservative because they are just above the naturally occurring conditions (90^{th} percentile of reference sites; Chapter 9). Setting conservative limits is consistent with DWQ's responsibility to prevent degradation of aquatic life uses. However, without including ecological responses, setting TN and TP limits this low could potentially result in impairment determinations for streams without obvious signs of biological degradation. Such a scenario was observed in this investigation where approximately two-thirds of the sites that were above the lower thresholds of

TN and TP did not show any signs of biological degradation. In addition, the use of ecological responses helped identify one stream with extremely high algae cover that would have been missed had DWQ relied exclusively on water column nutrients. This site was unusual because its growing-season average of TN and TP was below the lower thresholds for both TN and TP.

At first glance, the large proportion of streams (37%) considered to be impaired under the proposed headwater NNC seems high, because the study intentionally selected sites at which elevated nutrient concentrations had been observed during previous data collection efforts. While the growing-season average of TN was only slightly higher than the average among all previously sampled headwater streams, TP was 66% higher than the growing-season average of all the headwater streams evaluated (Chapter 9). As these proposed NNC are implemented, additional sites with nutrient-related impairments will be identified, but these results suggest the number of new problems will likely be small enough that they can be reasonably addressed with existing regulatory tools.

Conclusion

This study and the previous investigation to generate nutrient S-R relationships have clarified the effects of anthropogenic increases in N or P on Utah's stream ecosystems. The most general insight has been that it is often difficult to predict the response that an increase in nutrients can be expected to cause. Such variability among stream responses complicate the development of nutrient criteria, in Utah and elsewhere. The integration of ambient nutrient concentrations and ecological responses into the proposed NNC helps address this complication, because it allows the use of relatively conservative nutrient concentrations while minimizing the likelihood of incorrectly identifying impairments where none actually occur.

SECTION 4

Ongoing Implementation of Nutrient Reduction Programs

Chapter 14

ONGOING IMPLEMENTATION OF NUTRIENT REDUCTION PROGRAMS: AN OVERVIEW AND BACKGROUND

Key Points

The proposed headwater NNC require consideration of how these criteria will be incorporated into existing water quality programs.

Because interpretation of the proposed NNC requires collection of both ambient nutrient water chemistry and ecological responses, it was important to incorporate the collection of these data into existing monitoring efforts.

The proposed NNC include a basic assessment framework, additional analytical details with be provided in the analytical method of the Integrated Report.

While impairments are typically addressed with TMDLs; however, most human-caused sources of nutrients in headwater streams are from nonpoint sources, so alternative remediation approaches may be more appropriate for any impairments that are identified.

Introduction

The specific application of the nutrient thresholds and ecological responses to Utah Division of Water Quality (DWQ) water quality programs is of keen interest to stakeholders. Specifically, stakeholders are interested in how this information will be used to inform monitoring, assessments, numeric nutrient criteria (NNC) development, total maximum daily loads (TMDLs), and the development of permit limits. This chapter provides a broad overview of applications of the data presented in the first two sections of the report.

Monitoring and Assessment

Monitoring

Development of the S-R relationships discussed in this report involved repurposing existing monitoring data and developing new monitoring and analytical methods for several functional responses (see Appendix A for data acquisition standard operating procedures [SOPs]). The decision to augment DWQ's monitoring programs involved careful consideration of future monitoring logistics and the need to collect data for new indicators on a routine and ongoing basis. In addition to collecting data on the new indicators, DWQ will continue to independently monitor and assess several nutrient-related responses with existing numeric criteria (i.e., dissolved oxygen [DO] and pH), though these indicators were not studied in this technical support document (TSD). Detailed plans for integrating the new monitoring elements with data that have been historically collected will be included in DWQ's *Strategic Monitoring Plan*. This section presents a broad overview of how DWQ's monitoring will be carried out now that the set of data being collected has expanded.

Utah currently monitors six major basins using a tiered, rotating monitoring approach that combines the strengths of both systematic and random site selection. Randomly selected sites are used for routine assessment purposes and are conducted in year one of the rotation; systematic monitoring is used to support regulatory programs and is conducted in year three of the rotation. Each of the six major basins is visited in two different years within each rotation; all six basins are visited twice within a single six-year rotation, with two basins visited per year.

With respect to randomly selected moinotiring, 25 stream segments within the basin are selected using a stratified-random (generalized random tessellation stratified) draw from all perennial streams statewide. At each randomly selected site, biological assessments are conducted from summer through early fall (i.e., the growing season). Monitoring at these locations includes collections of water chemistry data, habitat descriptions, macroinvertebrates, diatoms, benthic algal abundance estimates, and data on observed fish species.

To this set of data that is already being collected, DWQ proposes to add the collection of filamentous algae cover. High-frequency DO and temperature for the purpose of obtaining metabolism indicators will be collected at enriched sites during the subsequent intensive basin monitoring cycle for the watershed where the stream resides. To collect the high-frequency DO and temperature data, sondes will be deployed during a visit to the location during the growing season and then retrievd the subsequent month yielding approximately 30-days of daily metabolism data.

Data from these indicators will be evaluated, and if sites are determined to be threatened by cultural eutrophication, more intensive monitoring can be conducted during the intensive, or programmatic, monitoring that occurs on year three of the rotating basin schedule. This will allow DWQ to obtain additional water chemistry samples. The intensive monitoring, which creates an opportunity to collect additional samples, expands the dataset from the 1–2 samples that are obtained from each stream site during the year-one monitoring effort. Other follow-up monitoring for the purposes of developing site-specific standards or TMDLs will also be incorporated into intensive monitoring schedules.

Nutrient-Specific Assessments

In accordance with U.S. Environmental Protection Agency (USEPA) integrated report guidance, DWQ's *Integrated Report* will propose nutrient-specific assessment methods. Once NNC are established for headwater streams, assessments will be conducted in accordance with the assessment objectives specified in the criteria and associated implementation materials. For streams lower in watersheds, nutrient assessments will be tied to Utah's narrative criteria. Assessing sites based on the narrative criteria is consistent with DWQ's ongoing biological assessment program and USEPA guidance. Further support is provided by a recently adopted clarification to Utah's narrative criteria, which requires that "waters of the state shall be free from human-induced stressors that will degrade the beneficial uses [of the waters]," and explicitly states that biological assessments can be used to determine whether uses are supported (Utah Administrative Code [UAC] R317-7.2)

DWQ plans to derive quantitative nutrient assessments from numeric indicators established from the S-R relationships described in this TSD; the indicators will be analyzed in conjunction with one another, not as isolated data points. DWQ currently plans to place sites with nutrient-related impairments on Utah's 303d list in a subclass of impaired waters. As with all impaired waters, a TMDL may be required for nutrient-related impairments. However, the intent of the subcategorization is to encourage the implementation of TMDL alternatives in appropriate circumstances. This is particularly relevant to habitat-limited sites that are in environments with multiple stressors. Under these circumstances, follow-up investigations will first emphasize the relative roles of all human stressors and, if appropriate, the best attainable conditions for the waterbody.

Specific assessment methods will be developed and published as part of the analytical methods in Utah's *Integrated Report*. The final assessment methods in the *Integrated Report* are sufficiently parsimonious that assessment decisions can be made consistently, following methods that can be easily communicated to stakeholders. In developing these methods, a risk-based approach was used that considers the magnitude of excursions above (or below) numeric indicators for both causal and response parameters. Assessments for parameters with existing numeric criteria, like pH and DO, will remain independent. In contrast, the functional indicators developed in this TSD will be interpreted collectively. These assessment procedures will also avoid the oversimplification that results from averaging or otherwise combining responses that measure different ecological processes. Just as no single indicator can conclusively identify a nutrient impairment, a combined score oversimplifies and eliminates many of the advantages of using multiple lines of evidence to more accurately quantify nutrient-related ecological responses.

Development of Numeric Nutrient Criteria

The S-R pilot study and analyses were primarily intended to be used in the development of water quality standards. The project gradually evolved as Utah's nutrient reduction strategy developed, but establishing NNC, where appropriate, remains a program priority. This section describes describes alternative approaches that DWQ could use to apply the S-R relationships presented in this TSD to derive NNC on an iterative basis for other streams in Utah.

Numeric Nutrient Criteria Development for Headwater Streams

The most immediate applications of the S-R relationships to NNC will be the adoption of the proposed NNC for headwater streams—those classified with Category 1 or 2 Antidegradation Protections (UAC R317-2)—into Utah's water quality standards. Chapter 10 provides several technical details and analyses in support of the proposed headwater NNC. These analyses include an evaluation of whether headwater streams require further classification and a comparison of the distribution of headwater N and P with those predicted to occur among all of Utah's streams. These analyses, together with multiple S-R thresholds, provide the information needed to develop headwater NNC.

A separate document, *Numeric Nitrogen and Phosphorus Criteria: Utah Headwater Streams* provides specifications for the NNC rule language, a summary of the underlying rationale for this proposal and programmatic considerations for implementation of the criteria. The decision to propose a combination of nutrients and ecological responses is predicated on the assumption that, while regional criteria are broadly protective of headwater streams, in some circumstances the physical characteristics of specific sites make the stream naturally resilient or sensitive to nutrient enrichment. In accordance with recent guidance (USEPA 2013) on combined criteria, this document also reviews how the combined criteria will apply to DWQ regulatory programs.

Site-Specific Numeric Nutrient Criteria Development for Streams Lower in Watersheds

The decision to prioritize establishment of NNC for headwater streams means that NNC for other streams and waterbodies will need to be developed on an ongoing, as-needed basis. Site prioritization of these streams and waterbodies will be based on a combination of data availability and regulatory needs. Approaches for site-specific criteria development will depend on factors such as the nature and extent of biological degradation, the potential for the effects of nutrients to be confounded with those of other human-caused stressors, and regulatory needs.

While specific approaches will differ based on the particular needs of the stream or waterbody, DWQ anticipates that investigations to support site-specific standards will generally follow procedures similar to those outlined in USEPA's Causal Analysis/Diagnosis Decision Information System (CADDIS) program (USEPA 2010), albeit with a modified objective. All nutrient-concentration data, ecological response data, and other human-caused stressor data will be compiled for the site. This evidence will then be organized into a conceptual model that describes all applicable pathways between nutrients and the ecological responses of interest. In planning stages, these models will primarily be used to identify pathways and responses of greatest interest and associated data gaps. The data review will then be used develop a sample analysis plan (SAP) that defines specific approaches with respect to data management, data acquisition, and proposed analytical approaches. As SAPs are developed, care should be taken to account for as many sources of potential uncertainty as possible, including the spatial and temporal variation of both causes and responses, and covariates with a strong likelihood of influencing causal inferences (see Chapter 11 for a more complete review). In most cases, the development and review of SAPs and subsequent project implementations will be a collaborative process involving DWQ, USEPA, and other engaged stakeholders.

Site-Specific Numeric Nutrient Criteria Development: Alternative Scenarios

Regionally derived S-R thresholds for nutrients and response parameters may need to be modified on a site-specific basis, particularly if specific responses are to be used to define water quality goals for a stream. Nevertheless, a careful review of causal and response parameters can be helpful in estimating the likely direction and scope of efforts to develop site-specific standards. To illustrate the utility of multiple lines of evidence, this section provides several alternative approaches for site-specific investigations based on observations of both causal and response parameters. These scenarios are not comprehensive, nor are the examples exhaustive, because each alternative involves many details that will need to be determined on a case-by-case basis.

Nutrients are High, but Ecological Responses Indicate Healthy Conditions

In some circumstances N <u>or</u> P may exceed regional numeric indicators even as structural <u>and</u> functional responses suggest that the biological integrity of the stream is high. Initially, site-specific NNC for these streams could be established from background conditions, under the regulatory authority of antidegradation provisions, to ensure ongoing protection from cultural eutrophication. Compared to other scenarios, this level of effort is relatively small, because once it has been established that existing conditions are fully protective, subsequent collection efforts can focus on ensuring that proposed NNC are derived from samples that represent the temporal variation of N or P at the site.

NNC that are established to be protective of current conditions do not necessarily preclude future proposals to increase nutrients in the stream. However, project proponents would need to demonstrate that the increases were necessary to accommodate "important economic or social development" and that the "instream water uses shall remain protective" (UAC R317-2-3.1). To demonstrate the latter the project proponent would need to provide justification that high site-specific N or P criteria would remain protective of the stream's existing uses.

Nutrients are High, and Ecological Responses Suggest Potential Degradation of Uses

Ongoing monitoring data may reveal that regionally derived numeric indicators for either N or P are high and one or more ecological responses suggest degradation of existing uses. When this occurs, site-specific investigations should first confirm that the responses are wholly or partially the result of excess nutrients. In these circumstances, SAPs may be developed that focus on the relative role of multiple stressors in degrading responses, or on the extent to which natural conditions (covariates) may be exacerbating deleterious responses.

Once the deleterious responses have been prioritized and associated thresholds confirmed or modified to accommodate naturally occurring site-specific conditions, the thresholds can be used to inform NNC and, to the extent that N or P is a causal parameter, TMDL endpoints. Meeting this objective may require development of a second SAP to obtain the data necessary to couple ecological responses with process-based (mechanistic) models.

DWQ and Utah State University recently completed an investigation that explored data requirements for using the combination of Qual2K models and field observations to generate site-specific NNC (Nielson et al. 2012). That investigation included recommended data quality objectives and development of an approach for collecting the requisite data for model calibration. That report, coupled with the data acquisition SOPs for the ecological responses described in this report, can be used as a starting point for the development of these SAPs. All the ecological response data and preliminary models are already available for most of the receiving waters of Utah's mechanical wastewater treatment facilities, so much of the preliminary data needed to inform the development of these site-specific investigations are already available. However, the specific data gaps identified for each site will need to be addressed by modifying collection methods for existing responses, developing alternative ways to summarize existing responses, or including novel ecological response data.

Ecological Responses Suggest Water Quality Problems, but Conditions Are Irreversible

Circumstances will arise when attainment of regionally derived numeric indicators for causal and response parameters is not feasible due to factors such as atypically natural conditions or irreversible habitat and hydrologic modification (40 CFR 131.10(g)). Under these conditions, the S-R relationships described in this report can be used to inform the development of specific causal or response water quality goals that best express the best attainable condition for the stream. To determine best attainable conditions for these streams, site-specific study designs will likely need to focus on the relative roles of nutrients and other stressors, mitigation efforts that can be feasibly implemented to restore all uses, and the causal and response parameters that are most likely to provide accurate and sensitive measures of stream conditions. In many cases, these investigations may reveal the need to modify the designated uses of the stream in conjunction with any proposed NNC, and the data obtained will provide the technical rationale for a Use Attainability Analysis that is required to make these changes to water quality standards.

Addressing Nutrient-Related Impairments

The S-R relationships developed in this report provide useful information to help DWQ and stakeholders more efficiently and effectively address streams with nutrient-related problems.

Watershed Prioritization

Nutrient sources are dispersed, the science required to understand site-specific responses to nutrient enrichment is complex, and fixing nutrient-related problems can be expensive. The consequence of these complexities is that DWQ cannot address all eutrophic watersheds simultaneously. Instead, a process for watershed prioritization is required.

DWQ, in partnership with USEPA and TetraTech, has been developing several prioritization tools. Results from an economic study have been georeferenced and incorporated into spreadsheet models that can be used to provide cost:benefit information on the economic consequences of NNC implementation on a watershed-by-watershed basis. The results generated by these economic models have been incorporated into a recovery potential screening (RPS) tool (Norton et al. 2009) that prioritizes watersheds based on the likelihood that sufficient nutrient reduction efforts—and other related restoration efforts—will restore or improve biological integrity. The RPS combines numerous GIS-based indicators to describe three indices that each measure different aspects of whether restorations are likely to succeed: a Social Index, an Ecological index, and a Stressor Index. The social index combines several metrics that evaluate sociopolitical aspects of restorations, such as the anticipated costs of remediation efforts or the commitment of local watershed groups to implementing restoration efforts. The other two components of the RPS tool are more easily related to the S-R relationships discussed in this report. The ecological index captures the extent of impairments (such as the departure from expected conditions) and directly aligns with multiple response parameters that collectively describe several causal pathways between nutrients and uses. The stressor index evaluates the number of stressors and the extent to which they can be improved to achieve water quality goals.

Among other things, the proposed site-specific approaches are intended to help address complications arising in multiple-stressor environments. Alignment of the RPS tool with the overall nutrient reduction strategy provides a way to estimate the relative cost and complexity of site-specific criteria development for different watersheds. More importantly, this tool may provide insight into places where these resources expenditures are most likely to result in improvements to water quality.

Alternatives to the Traditional Total Maximum Daily Load Approach

Integral to Utah's nutrient reduction strategy is an action-oriented approach for addressing nutrient-related water quality problems. It is likely that DWQ will identify sites that have nutrient-related problems but do not yet have NNC. Development of NNC will take time; in the interim, an approach is needed that will allow known nutrient sources—particularly those that are relatively inexpensive to fix—to be addressed on an ongoing basis.

Historically, water quality impairments have been addressed with TMDLs on a pollutant-by-pollutant basis. TMDLs remain an integral part of DWQ's water quality management strategy; however, TMDLs have several drawbacks in addressing N and P pollution. First, TMDLs work best in situations where all sources can be clearly demarcated and quantified, which will be especially difficult with nutrient pollution given the large number of both natural- and human-caused nutrient sources. Second, nutrient pollution is often found in conjunction with other causes of degradation, in which case considering each pollutant individually may not lead to the most efficient or effective remediation practices. Third, in some circumstances, it can be more cost effective to remediate the effects of nutrients than to exclusively seek reduced N or P loads. For example, if a stream is functionally impaired due to excess primary production, it may be possible to mitigate these problems by restoring riparian ecosystems to increase channel shading. Riparian restoration efforts often have other desirable outcomes (i.e., aesthetics, erosion reductions) for recreation and aquatic life uses.

DWQ proposes that TMDL alternatives be considered as potential mechanisms for more efficiently and effectively mitigating nutrient-related impairments. An alternative process is proposed to: (1) identify all potential sources of nutrients or other stressors that may be contributing to the degradation of structural or functional responses, (2) convene appropriate stakeholders, and (3) develop and implement incremental, watershed-specific restoration efforts with the goal of restoring ecological responses. This proposed program is action oriented, because it allows for the most cost-effective restoration efforts to begin quickly while the science necessary to establish site-specific NNC—or TMDL endpoints—is ongoing. The program is also cooperative, because it potentially allows stakeholder to come together to find local solutions to common goals. The proposed program also requires accountability because ongoing monitoring would be required to demonstrate iterative improvements, either in direct N or P reductions or in measures of functional or structural condition. Demonstration of iterative progress would be more likely because different indicators respond to restoration efforts at different spatial and temporal scales. This proposed approach is not intended to entirely replace TMDL requirements, but structured endpoints and load allocations may not be needed if alternative methods can demonstrate iterative progress toward meeting water quality objectives.

Process-Based Models: Support of Permit Limits and Site-Specific Standards

A collaborative study between DWQ and Utah State University was initiated in concert with the S-R investigations with the dual purpose of providing tools for site-specific NNC development and improving the accuracy of Wasteload Analyses and associated permit limits. That study generated several important products that will help integrate the S-R information with process-based models as ongoing elements of Utah's nutrient reduction strategy.

The study developed methods for data acquisition procedures to populate and calibrate QUAL2Kw models. It also developed a method for consistently and repeatedly using the data to calibrate the models. The study included a sensitivity analysis that identified the model parameters most critical to the accuracy of model predictions. The results of the investigation are provided in Chapter 12 and Appendix C.

Chapter 15

ACCOUNTING FOR COMPLEXITY, UNCERTAINTY, VARIABILITY, AND COVARIABLES IN SITE-SPECIFIC ANALYSES: THE PATH FORWARD

Authors: T. Miller, D. Richards, and B. Marshall

Key Points

This chapter reviews important sources of variability and complexity in aquatic ecosystems, both within natural stream gradients and in response to human-induced stressors.

Ecological processes are naturally variable through space and time, which can lead to uncertainty and errors in data interpretation.

Inaccurate interpretations and assessment errors can be minimized with the design and implementation of robust sample designs.

Appropriates statistical analysis can also minimize inaccurate interpretations of field observations, but statistical significance should not be confused with ecological significance.

Incorporation of important covariates when evaluating stressor-response relationships will improve the accuracy of site-specific numeric criteria as they are developed in the future.

Introduction

Acute and chronic thresholds for known toxicants to aquatic life were established in relatively short order by EPA and with relatively few revisions (e.g., in the Environmental Protection Agency's [EPA's] "Red Book" [1976] and "Gold Book" [EPA 1986]). These criteria have been a major step toward maintaining and restoring healthy aquatic ecosystems. One of the greatest remaining challenges in meeting Clean Water Act (CWA) goals for "maintaining and improving the physical, chemical, and biological integrity of our nations waters" and "protecting and restoring the nations waters to be swimmable and fishable" is controlling anthropogenic eutrophication, which is one of the most common topics of contemporary applied aquatic ecology literature. Researchers and managers face many obstacles in establishing water quality criteria for nutrients. For example, other than rare cases of high concentrations of ammonia or very high concentrations of nitrate, many forms of nutrients are not toxic to aquatic life or human health. Rather, significant changes to important structural and functional characteristics of aquatic ecosystems can occur at concentrations much lower than toxic levels.

One of the earliest accounts linking nutrient enrichment and algal blooms was documented by Edmondson et al. (1956), and key seminal papers were published in the 1960s (e.g., Vollenweider 1968) and 1970s (e.g., Schindler et al. 1971). The link between phosphorus (P) and phytoplankton was well established for freshwater lakes and worldwide, and P- reduction programs began in the 1970s. The "P-limitation" paradigm continues today (Smith and Schindler 2009), although notable exceptions have been reported, including ephemeral or seasonal nitrogen (N) limitations in lakes and estuaries (e.g., Howarth and Marino 2006, Nydick et al. 2004) and N-limitation and N and P co-limitation for some streams (Dodds et al. 2002).

Algal or microbial taxa acquire and utilize N and P in different ways and early observations indicated that environmental variable interactions often produced taxa- specific outcomes (e.g., Mitsui et al. 1986, Tank and Dodds 2003). In addition, lotic ecosystems respond spatially and temporally to many factorsincluding: sediment loads, hydrology, geomorphology, , temperature, discharge, velocity, substrata particle size, (Burdon et al. 2013, Dewson et al. 2007, Hynes 1970), riparian condition/density, (e.g., Hynes, 1970, Maloney and Weller 2011), detritus sources (Dodds 2007); watershed size (Heathwaite 2010), and land-use (Allan 2004, Townsend et al. 2008). Many of these changes occur as natural gradients as streams flow from headwaters to downstream termini. These natural gradients of physical characteristics and concomitant biological responses result in ecosystem transitions that have been defined within the context of the River Continuum Concept (Vannote et al. 1980). When human activities influence these variables they are labeled "anthropogenic" or "human-induced" stressors and when overlaid natural gradients can create significant challenges for scientists and managers when sorting through these covariables to identify and prioritize causal factors at impacted (impaired) streams.

The purpose of this chapter is to identify important sources of variability and complexity in aquatic ecosystems, both within natural stream gradients and in response to human-induced stressors. We discuss strategies to account for variability and complexity when translating regional-based numeric nutrient criteria (NNC) to the development of site-specific NNC; these strategies account for natural stream transitional zones as well as multiple human-induced stressors affecting reaches that appear to be nutrient enriched. In the context of specific management objectives, the chapter provides suggestions for appropriately adjusting the relative weight of ecological responses and regional field and laboratory methods to account for locally important sources of natural variation. Where significant land use change and hydrologic modifications have occurred, more intensive investigation is often required to evaluate the relative importance of nutrients versus other stressors. Identifying key stressors allows the prioritization of management strategies best suited for immediate restoration efforts and can vastly increase the chances for success, both ecologically and socially.

Complexity, Uncertainty, Variability, and Covariates

Aquatic ecosystems are known for their inherent complexity. All analyses of stream ecosystems are faced with elements of uncertainty, and stream ecosystems are usually observed incompletely through sampling, resulting in varying levels of uncertainty (Frey 1993). Seife (2011) captures this concern in the statement that "any attempt to measure something is prone to error." The expense of sampling often prohibits collecting as much data as needed to account for and reduce uncertainty (Frey 1993). Although data may be limited to support regulatory assessments, criteria development, or site-specific analyses: decisions must be made. Therefore, the objective of resource managers is to determine the "maximum acceptable" level of uncertainty in criteria development. Such decisions require that important sources of uncertainty, particularly those associated with specific management objectives, are acknowledged and every effort made to reduce these sources to the greatest extent possible.

Ecosystems are naturally variable, both spatially and temporally (Townsend 1989, Ascough et al. 2008). Virtually every biological and ecological process is variable, consequently variability is a common source of uncertainty in assessments. The distinction between variability and uncertainty is not always clear and may be context dependent, particularly when only a limited amount of empirical information is available (Frey 1993, Hayes et al. 2006). Most assessments must deal with both uncertainty and variability simultaneously and variability is often used to assess certain types of uncertainty. Although it is tempting to ignore uncertainty and variability because they can make assessments and the development of site-specific criteria challenging, they must be accounted for in the development of water quality management programs. This chapter aims to provide guidance to DWQ in accounting for uncertainty when developing NNC.

Addressing Uncertainty and Variability

In general, there are three main types of uncertainty: linguistic uncertainty, epistemic uncertainty, and variability. It is important to differentiate these types of uncertainty before integration and development of site-specific criteria (Ferson 1996, Ferson and Ginzburg 1996, Regan et al. 2002, others).

Linguistic uncertainty comes from the difficulty in communicating precise meanings. Vagueness, ambiguity, and context-dependence are sources of linguistic uncertainty (Gregory et al. 2012). Context is everything. Assessments and site-specific analysis can reduce linguistic uncertainty by clearly stating and defining important concepts and terms during their development (Gregory et al. 2012, Regan et al. 2002).

Epistemic uncertainty is uncertainty associated with knowledge or lack of knowledge of the state of an ecosystem (Hayes et al. 2006, Morgan and Henrion 1990) and pervades all our attempts to discover the truth about ecosystems and our ability to make sound assessments or to develop meaningful management criteria (Ascough et al. 2008). Uncertainty can be reduced by additional research—the parameter value can be refined and then further quantified (Frey 1993, Hayes et al. 2006). For example, site-specific estimates of P- levels that result in blooms of two nuisance algal taxa, *Cladophora* sp. and *Didymosphenia* sp., can be further refined by additional research.

There are many types of epistemic uncertainty, including measurement error, data uncertainty, systematic error, statistical uncertainty, model uncertainty, parameter uncertainty, inherent randomness, and subjective judgment (Frey 1993, Hayes et al. 2006). Data uncertainty from measurement error can occur throughout the assessment process and can often be confounded with variability. Systematic error is the error that is constant in repetitions of the same experiment, observation, or sampling protocol. An example of systematic error is continually assigning a macroinvertebrate species to a functional feeding group based on family-level functional feeding group classification (or even misclassification) instead of using a species- or genus-level assignment or by reclassifying the species based on additional literature review. For example, the chironomid midge subfamily Tanypodinae are generally considered predatory, but *Tanypus neopunctipennis* is a collector-gatherer that can be a facultative scraper of epiphytic diatoms (Marshall, unpublished data). The role of this species in models is highly dependent on taxonomy, and the subfamily-level functional feeding groups would misclassify these highly productive species in functional modeling efforts.

> "No amount of statistical sophistication can rescue a poorly designed study." (Lovell 2013)

The opposite of systematic error is random error. Systematic errors are usually much more serious than random errors because their magnitude cannot be reduced by simple repetition, and these errors often go unrecognized. Additive systematic error is known as 'bias.' The only way to deal with systematic error is to recognize a bias and remove it by thorough examination and validation of the assessment procedure or control it with a stratified sampling design.

Sampling Design

The first step toward reducing error and uncertainty is to articulate a clear research question and accept that spatial and temporal environmental variation exists. The next step is to develop an appropriate sample design that is relevant to the research question and accounts for inherent spatial and temporal variation. Appropriate sampling design and data collection are necessary prerequisites for developing robust site-specific criteria regardless of how the data are analyzed. Stratification, randomization, and replication are imperative in all site-specific sample designs because they reduce error and uncertainty (see section below, "The Path Forward" and for detailed descriptions of improved macroinvertebrate sampling design).

Power analyses can help determine the ideal number of samples per stratum to balance sensitivity and cost of the assessment. Although randomization is important, randomization alone is inadequate. A completely randomized design without stratification assumes either that the study subject occurs completely randomly in the study area or that the researcher has no *a priori* knowledge of the subject. Nothing in nature is random, and ecologists have amassed a great deal of information about the spatial and temporal distributions of most organisms in streams; therefore, stratification is imperative. Stratification can help focus assessments on specific stressors by reducing variability associated with generalized random sampling designs. Sampling design strategies for site-specific assessments can be found in Thompson and Seber (1996) and Chao and Thompson (2001). Site-specific sampling methods must also complement the sample design and will likely differ from those used for region-wide sampling methods. Knowledge and avoidance of pseudoreplication (Hurlbert 1984) are required in all site-specific sampling designs but is not usually in the form that most biologists conceive of; samples should be appropriate for the specific hypothesis or research question addressed by the study. Thus regional-scale replication is not appropriate for site-specific criteria.

Given that many ecological processes evolve dynamically, both spatially and temporally, purely spatial sample designs are often not as efficient as those that consider spatiotemporal dependence (see Underwood 1996). Many methods and techniques are available that model both factors simultaneously: hierarchical models such as network designs (Le and Zidek 1994), optimal designs for time-dependent responses (Federov and Nachtsheim 1995), models for updating sample design in repeated environmental surveys (Arbia and Lafratta 1997), optimal network designs for spatial prediction covariance parameter estimation, empirical prediction (Zimmerman 2006), and dynamic design networks for space-time models and non-Gaussian data (Wikle and Royle 1999, 2005).

Statistical Approaches

Frequentist statistics are the most well-established and commonly used methods of statistical hypothesis testing (i.e., null hypothesis significance testing [NHST]). However, NHST has come under criticism of late (Goodman 2008, Johnson 1999, Seife 2010, Ziliak and McClosky 2009). Bayesian statistics provide an alternative framework to that provided by frequentist statistics. Both schools of thought have their place and usefulness in ecological assessments, and a thorough understanding of their nuances and applicability are needed. When Frequentist NHST methods are considered for use, understanding their limitations and the appropriate use of Type I error (false positives), Type II error (false negative), and the meaning of p-values is necessary.

Statistical methods that don't rely on NHST and p-values that could be considered and evaluated for site-specific criteria include estimation approaches based on confidence limits, maximum likelihood, or quantile regression; multivariate methods such as hierarchical clustering, non-metric multidimensional scaling, classification and regression trees; Akaike information criterion for model evaluation; structural equation models; non-parametric multiplicative regression; or fundamentally different approaches based on Bayesian statistics. In addition, there is increased promise of incorporating machine learning methods such as Random Forest models for large datasets and the use of meta-analysis for statistically comparing results of many studies. The selection of the most appropriate method(s) will depend on the specific questions to be addressed, the level of uncertainty that is acceptable during NNC development, and the parameters and data chosen for evaluation.

Several of these multi-metric methods are mostly exploratory in nature, such as hierarchical clustering and non-metric multidimensional scaling, whereas Random Forest models, structural equation models, and non-parametric multiplicative regression are more confirmatory. Bayesian methods can be both exploratory and confirmatory but allow *a priori* estimates and information to be included in the analyses. Multiple lines of evidence are also quite useful and can add strong decision support, particularly if they measure several different ecological responses. As previously mentioned, the chemical, physical, and biological responses to excess nutrients are varied and complex. Multiple lines of evidence that describe different aspects of these responses can be pieced together to paint a more complete picture of nutrient-related problems or other covariables that are causing the observed degradation. Such knowledge is critically important to minimizing assessment errors. More importantly, a more complete understanding of the problem frequently informs the selection of restoration activities that are most likely to efficiently and effectively achieve water quality objectives.

Ensuring Biological Significance

Statistical significance and biological significance are usually not the same. Statistical significance is almost meaningless in site-specific assessments if it is not biologically relevant. Defining what is biologically significant or relevant is not straightforward, nor is it a statistical decision. Defining biological significance has important consequences for the design, statistical analysis, and interpretation of a study and development of site-specific criteria (Lovell 2013, Martinez-Abrain 2008). Biologically significant effects can be quantified by determining "effect size" and using power and sample size calculations (Lovell 2013, Martinez-Abrain 2008). The choice of an appropriate effect size requires a thorough understanding of the biological subject and the context in which it is being evaluated Expert judgment and knowledge of the subject are required, and there may be no consensus on the minimum difference (effect size) considered significant (Lovell 2013, McBride and Burgman 2012). For instance, Cohen (1988) considers 20% a statistically small effect-size, but the loss of 20% of species, may not be an acceptable management objective. The choice of the effect size is, therefore, a decision for the subject matter expert(s) and should be based on knowledge of the topic. In the case of site-specific criteria development, managers are responsible for deciding on the effect size (Lovell 2013).

Power analysis is a useful tool for informing biological relevance (i.e., level of difference, effect size) because it captures variation in the context of effect size. Often the magnitude of ecological response is a more appropriate decision tool than NHST p-values, which only identify whether measures of central tendency differ among populations. In the case of power analyses, p-values can be replaced by confidence intervals (CIs). Since effect size is decided by the subject matter expert(s), the choice of upper and lower CI bounds or CI intervals is not a statistical decision (Lovell 2013). The balance of environmental sensitivity with statistical sensitivity can be evaluated using power analysis to evaluate the ratio of statistical effect size (δ) with field observations of change (Δ) as a metric of sensitivity (Marshall 2001).

> "Remember that all models are wrong; the practical question is how wrong they have to be to not be useful." Box and Draper (1987)

Model Uncertainty

Models are simplified representations of a real-world system and therefore are likely to be incomplete and are often a key source of uncertainty (Frey 1993, Hayes et al. 2006). Alternative models are often available, and the choice and use of the most appropriate model(s) can reduce uncertainty considerably. Each model comes with its own suite of input parameters and decision rules. For example, most multivariate methods provide the researcher with a plethora of statistical model choices (e.g., type of distance matrix, maximum likelihood vs. bootstrapping, or relying on default values, etc.). Assumptions need to be made throughout the modeling process. Many of these choices may be equally valid or may not have been thoroughly evaluated for use in ecological data and their selection can significantly affect the outcome of the analysis. In addition, the results of these choices are multiplicative and can result in numerous combinations of models, each of which can result in differing model outcomes. Uncertainty imposed by choice of models and the cumulative effects of input parameters and decisions rules needs to be carefully examined and addressed prior to analyses.

Increased model complexity does not always reduce the amount of uncertainty (Frey 1993; Hayes et al. 2006). An increased number of parameters in a model may, in fact, increase the uncertainty of the model outcome for a given set of data and can be either additive or multiplicative or both (McCune 2011)—especially when some model parameters themselves are derived from other models (e.g., PRISM climate models). When there is more than one uncertain quantity, uncertainties may be statistically or functionally dependent or correlated (Cressie et al. 2009). Failure to properly model the dependence between the quantities can lead to uncertainty in the result, primarily in the variance of output variables.

The most useful models will provide the greatest simplifications while providing an adequately accurate representation of the processes affecting the phenomena of interest (Walker et al. 2003). Merow et al. (2014) suggested that researchers should constrain the complexity of their models based on study objective, attributes of the data, and an understanding of how these interact with the underlying biological processes. The purpose of a quantitative uncertainty analysis is to quantify the degree of confidence in an analysis or assessment using the most appropriate data and models available (Ferson and Ginzburg 1996, Ferson et al. 1999). Sensitivity analysis can be used to determine where uncertainty reduction is most necessary and beneficial. The use of probability and interval analysis can deal with all epistemic uncertainty (Ferson and Ginzburg 1996, Ferson et al. 1999); however, the increased accuracy of probability distributions and interval analysis can result in decreased precision (Richards 2009). Accuracy is always more important than precision (Ziliak and McCloskey 2009); if a qualitative assessment is being conducted, precision may become unnecessary or non-essential (Dambacher et al. 2003).

Technical Uncertainty

In addition to the types of data uncertainty discussed, there can also be technical uncertainty. Technical uncertainty is the uncertainty generated by software or hardware errors— hidden flaws in the technical equipment (Frey 1993, Hayes et al. 2006). Software errors arise from bugs in software, design errors in algorithms, and typing errors in model source code (Walker et al. 2003). Many new, exciting, and potentially useful statistical methods and ecological models are published almost daily in well known, peer-reviewed journals or have been developed for regulatory agencies or commercial use. These methods should be used with caution while they are in their infancy and until they have been further tested and verified before incorporating into NNC. However, if carefully planned and evaluated, existing methods should be sufficient to develop meaningful site-specific criteria through iterative feedback and adaptive management.

Inherent Randomness

Some quantities may be irreducibly random; however, this concept is often applied to quantities that can be measured precisely but as a practical matter are not. In ecology, true randomness is rare, and inherent randomness is unlikely. Some processes that resemble randomness are actually the product of unmeasured variables or covariables; for example, the distribution of macroinvertebrate taxa within riffle habitat. Several strategies that can help reduce apparent randomness are discussed later in this chapter.

Model Output

This is the accumulated uncertainty that is propagated through a model. Model output uncertainty is the result of all the types of uncertainties listed above and variability, which is discussed below. Model output uncertainty is often ignored or misunderstood. One often overlooked outcome of accumulated uncertainty occurs when the total model error rate is larger than the predetermined significance level of the model results. Model results are likely invalid when this occurs (Seife 2010). Model output error is sometimes called prediction error. Uncertainty occurs at every stage and can accumulate throughout the analysis process. Uncertainty should always be accounted for by thoroughly examining every stage of the analysis and reducing it when possible and by reporting results with bounds such as probabilities and CIs; only then can we be confident in the scientific inferences made from an analysis (see Cressie et al. 2009).

Variability

> The true nature of rivers and streams: "Patchy in space, dynamic in time." (Townsend and Hildrew 1994)

Variability is a type of uncertainty that is also referred to as external, objective, random, stochastic, or natural variability It is related to the inherent variability in natural and human-altered ecosystems. Understanding natural variability is critical in management decisions since it is usually poorly understood and often confused with knowledge uncertainty (epistemic uncertainty; Frey 1993, Hayes et al. 2006).

Variance is a measure of the heterogeneity of an ecosystem parameter. Variance cannot be reduced by further research, but it can often be represented more accurately and communicated better with additional data. Variance is often best quantified as a frequency rather than a probability distribution (Frey 1993, Hayes et al. 2006). Additional mathematical techniques that can be used to address model uncertainty and variability include: qualitative modeling (Dambacher et al. 2002), Bayesian belief networks (Henrion et al. 2001), AIC, second-order Monte Carlo simulation (Cullen and Frey 1999), probability bounds analysis (Ferson 2002), information-gap theory (Regan et al. 2005), and hierarchical Bayesian techniques (Link et al. 2002).

Incorporating Covariates into Site-Specific Numeric Nutrient Criteria Development

In the context of this chapter, covariates include the sum of additional physical, chemical, or biological attributes that may operate as natural gradients or as human-induced stressors that confound interpretations of the extent to which the stressor of interest (i.e., nutrients) causes changes in ecological responses. Many of these characteristics have been described in the context of natural gradients of variability as streams flow from headwaters to larger streams with greater flows, warmer temperatures, lower gradients, etc. One of the earliest treatises on this subject was by Hynes (1970), who recognized the unique tolerances of specific periphyton and macroinvertebrate taxa to natural physical gradients and variables, including those that tolerated wide ranges of variability (tolerant taxa) and those that tolerated narrower ranges (intolerant taxa). Hynes (1970) characterized these natural gradients in physical and chemical characteristics as natural and predictable zones that favor specific taxa.

A decade later, Vannote et al. (1980) recognized that the transition between such zones occurs on a continual gradient and proposed the River Continuum Concept (RCC) as a more holistic view of watersheds where the gradient elicits a series of predictable responses within biological communities as populations follow a similar continuum of abiotic factors, energy sources, and the transport, utilization, and storage of organic matter and nutrients along the length of a river (i.e., nutrient spiraling [Webster et al. 1975]). Within the RCC framework, streams will generally transition from more oligotrophic allochthonous primary production in headwaters to a downstream mesotrophic state that includes an increase in the contribution and importance of autochthonous organic matter and nutrient loading from more developed riparian communities and larger watersheds (Dodds 2007, Vannote et al. 1980). This complexity is further complicated by seasonal transitions from primarily heterotrophy to autochthonous primary production (spring and summer) because of increased light availability and a reduction in allochthonous detritus abundance, for example, after leaf fall (Dodds 2007). As streams increase in size, there is usually a transition to lower physical gradients and smaller inorganic substrata size. In turn, this increases the ability of streams to retain, process, and store organic matter and nutrients within depositional sediments. In response, biotic communities also transition to include taxa that can occupy different or smaller substrates to capitalize on energy (organic carbon) that originated in upstream communities. This often leads to a dominance of heterotrophic-oriented taxa. These factors concurrently and continually interact to influence the habitat availability for specific taxa and the community assemblages of primary and secondary producers and macroinvertebrate populations that respond to these physical and biological patterns.

The importance of systematic changes along natural environmental gradients, like those described by the RCC, was acknowledged by EPA in their Nutrient Criteria Guidance: "A directly prescriptive approach to nutrient criteria development is not appropriate due to regional differences that exist and the lack of a clear technical understanding of the relationship between nutrients, algal growth, and other factors (e.g., flow, light, substrata)" (USEPA 2000). This statement only acknowledges "regional differences" in important variables, while the examples provided (e.g., flow, light, substrata) vary on a local site-specific basis. EPA acknowledges the importance of site-specific conditions in its water quality standards regulations: "EPA recommends that States and Tribes establish numerical criteria based on section 304(a) guidance, modified to reflect site-specific conditions, or other scientifically defensible methods."(USEPA 2000).

A much greater challenge lies in the fact that human activities have altered each one of these natural gradients by: (1) dewatering, (2) channelizing, (3) altering substrate composition, (4) altering riparian vegetation, or (5) altering natural landscapes in the watersheds. The cumulative result has been to exceed the typical ranges of normal variability for virtually all abiotic factors discussed above. The degrees of alteration of these factors may be additive or multiplicative in their effects on the structure and function of biotic communities (Folt et al. 1999, Underwood, 1989).

Secondary influences can also arise from the interactions of potentially new combinations of species in anthropogenically altered states that may result in "novel" ecosystems or simplified regimes (Hobbs et al. 2006) that have yet to be understood and described. Meaningful nutrient criteria will have to address this reality by working within an adaptive management framework, and future criteria modifications are likely.

The most common approach to understanding changes in biological communities has been to establish a suite of reference sites that represent a regional average of geographical and meteorological conditions and an abbreviated taxa list that most commonly occurs among these reference sites. Subsequent sample collection of the local taxa (e.g., macroinvertebrates) from a site that falls within that ecoregion is then compared to the list and is understood to be a fraction of the species on the regional reference list.

One key problem with using regional indicators in establishing reference condition is the lack of resolution and understanding of conditions that occur within major transitional zones of the river continuum. This ignores natural gradients in stream abiotic and biotic conditions, which naturally dictate species distributions. It would therefore be prudent to refine reference conditions by using higher spatial resolution to identify and stratify subsets of regional reference sites that represent the appropriate zones according to river continuum principles for that region. Ideally, these zones would incorporate an understanding of energy sources and flow within these zones. Accordingly, Dodds (2007) discussed the importance of primarily heterotrophic or autotrophic states, as observed within zones, in influencing nutrient utilization and, therefore, their influence on different thresholds of impacts. Even within such zones, site-specific conditions greatly influence stream metabolism and algal growth (Dodds 2007, Hill et al. 2009), and other characteristics. For example, flood frequency influences periphyton growth and accumulation and, all else being equal, eutrophication effects will likely be stronger under stable flow regimes (i.e., intermediate disturbance hypothesis; Biggs, 2000).

The Path Forward

When making determinations about the size and scope of site-specific investigations, numerous factors should be considered. An important first step in making these determinations is a careful examination of all water quality data to determine the strength and weaknesses of existing data. As an example, if the application of site-specific nutrient criteria would result in expensive wastewater treatment facility upgrades or best management practice implementations, then additional monitoring and research might be warranted to reduce the risks and costs of overly stringent criteria to local communities. On the other hand, the possibility also exists for extended deleterious impacts to streams while investigations are ongoing. Similarly, delayed regulatory decisions may result in additional legal vulnerability to DWQ or the regulated community. Taken together, the pros and cons of these decisions emphasize the importance of the perspectives of all stakeholders and the need for trust, collaboration, and consensus in reaching management objectives in both NNC development and implementation.

Further reduction in risk and uncertainty can be accomplished within the framework of adaptive management. Adaptive management is embraced by DWQ because it intrinsically acknowledges uncertainty by incorporating continued periodic monitoring that can provide data to be used in the refinement of management decisions. DWQ's approach to the development of NNC is an example of an adaptive decision that is intended to facilitate adoption of regional NNC for headwater streams while more nuanced issues of covariates and multiple-stressor environments continue to be addressed on a site-specific basis elsewhere. The decision to implement a statewide technology-based effluent limit for P of 1 mg/L is another important example of implementing an adaptive management strategy. In this case, DWQ and stakeholders acknowledged that most wastewater treatment facilities and industrial discharges add enough P to pose a threat to the integrity of receiving waters. DWQ and stakeholders agreed that the necessary treatment-facility modifications were not economically prohibitive. These modifications will provide for an initial and substantial (approximately two-thirds) reduction in effluent P concentrations and are expected to improve the biological conditions of many receiving waters. This will nearly halt additional P loading from point sources in the near future while additional studies can be performed on priority streams to identify where additional nutrient reduction is necessary or where other stressor(s) need to be mitigated to maintain or restore the biological integrity of receiving waters. The following section incorporates topics highlighted earlier into site-specific investigations as employed through an adaptive management framework.

Development of Study Designs

Once existing data available for a site has been compiled and placed into a conceptual framework, the next step is designing a study to meet management objectives. As previously highlighted in this chapter, this step is the most important if the study results are ultimately going to account for complexity and minimize uncertainty and variability. This section discusses general considerations for appropriate site-specific designs and provides specific examples of how these concepts might be applied to two interrelated questions important to the nutrient-reduction strategy: the site-specific modification of nutrient indicators (responses) and the development of site-specific numeric criteria.

General Considerations

Many of DWQ's routine monitoring and assessment procedures were developed to provide insight into regional conditions and water quality concerns. As site-specific designs are established, there are several contexts for which routine monitoring methods should be reconsidered to better match the specific study objectives at smaller spatial scales. This requires managers to carefully consider the scale-appropriate site selection, sampling regime (Price 1991), laboratory methods, and relevant ancillary data collection.

The Importance of Replication

Elements of replication are, of course, required to account for within-site variation and should be applied to all site measurements whenever possible. One way to determine the amount of replication required is power analysis. Wherever possible, replication should be done to account for each of the covariables addressed through a stratified sample design, which might involve both regional and site-specific characteristics. But in some cases, the current regional scale methods prevent use of replicated site-specific methods. In these cases it is especially important to ensure that the methods are suitable for the goal. If the goal is to develop site-specific criteria, but investigators are mildly curious about how it holds up to predeveloped regional criteria, they can destroy the study by using regional methods. For example, regional methods may use composited samples, these are very impractical to replicate, and impede the use of covariates.

Site Selection

Most study designs will require comparing data collected at the site of immediate interest to data collected at reference or control sites. If this is an element of the study design, care must be taken to ensure that these sites inform, rather than confound, management objectives. Regardless of the purpose, all sites should be selected to account for natural gradients that occur within single watersheds. Of particular importance are natural gradients in temperature, substrata, nutrients, allochthonous and autochthonous sources of organic carbon, and the extent to which natural changes in biological composition might be exerting top-down controls on stream biota.

In some circumstances, indicators will need to benchmarked against reference sites to derive estimates of background conditions. When this is necessary care should be taken to ensure that sites are selected to account for important covariables that are known to influence ecological responses. In Utah, most second- or third-order streams have been fundamentally altered after flowing from the relatively protected headwaters of U.S. Department of Agriculture Forest Service (USDAFS) lands, flowing through altered landscapes (agricultural or urban), experiencing dewatering for irrigation, and/or experiencing the effects of degraded riparian communities. Therefore, reference sites need to be identified that have minimal human disturbances but that otherwise have physical and chemical characteristics comparable to the site under investigation. Selection of appropriate sites will likely involve consideration of finer-scale attributes than DWQ typically uses for regional assessments because regionally important covariables may not be the primary determinate of biological responses at finer, local scales. This is particularly true for major transitional zones along the river continuum. For example, few Plecoptera in the Northern and Middle Rocky Mountain Ecoregions occur where summer water-temperatures exceed 20^{0}C (reference) or where sediment embeddedness exceeds about 40% and small gravel, sand, or silt particles dominate the substratum (Richards unpublished data). The absence of these taxa in mid- or low-elevation streams may not be due to nutrient gradients but rather to constraints by the physical conditions to which they are adapted. It would therefore be prudent to refine reference conditions by identifying and stratifying subsets of regional reference sites that represent the appropriate zones according to RCC principles for the region. This is necessary because other aspects of landscape changes may have equal or greater effects on invertebrates and periphyton assemblages than nutrients alone (e.g., Burdon et al. 2013, Maloney and Weller. 2011, Wagenhoff et al. 2012, etc.). Identifying the mechanisms behind these influences is complicated by the many potential pathways (often indirect) between land use and ecosystems and by the long-lasting effects of past land use. These indirect and lasting effects need to be better understood in order to support ecosystem restoration and conservation efforts (Maloney and Weller 2011).

Although the use of reference sites is preferred, control sites may be more appropriate and may better distinguish the relative roles of multiple stressors on ecological responses. Downes (2010) suggested avoiding the idea of reference sites altogether and using control-based site evaluations. Downes (2010) and Quinn and Keough (2002) suggest that the purpose of controls is to isolate the effect of a particular "treatment" (i.e., stressor). This means that controls must be as similar to the stressed sites as possible, except for the stressor of interest (i.e., nutrients). If stressed sites have a suite of other stressors present, then so should the control sites. Other sampling designs can also be considered if they can reduce uncertainty and account for variability and covariates better than before, after, control, and impact study designs (BACI designs).

Accounting for Ecosystem Processes

Ideally, site selection should also consider abiotic and biotic mediation of energy sources and flow. Dodds (2007) discussed the importance of trophic states, which are primarily heterotrophic or autotrophic, as influences on nutrient utilization—and therefore as influences on different thresholds of impacts. Even within such zones, site-specific conditions greatly influence stream metabolism and algal growth (Dodds

2007, Hill et al. 2009) and other characteristics. For example, flood frequency also influences periphyton growth and accumulation and eutrophication effects will likely be stronger under stable flow regimes than in systems subjected to frequent flow spates (i.e., intermediate disturbance hypothesis; Biggs, 2000).

Modifying and Expanding Ecological Responses

Several emergent structural and ecological response indicators, such as primary production and, stream metabolism and respiration, have been evaluated on a regional basis to develop response thresholds that can be used to improve the accuracy of nutrient-related assessments and provide information that can inform subsequent site-specific assessments. All regional indicators and responses, especially biological data, need to be evaluated at the site-specific scale(s). The DWQ headwater stream designated beneficial use is to protect cold-water fisheries and the aquatic food chain they depend on. The Clean Water Act also includes protection and propagation of fisheries and shell fisheries (i.e. mollusks: mussels, clams, and snails) and contains a critically important 'biological integrity' clause. Thus, there is a real need to primarily focus on fish, macroinvertebrate, and diatom metrics that are directly affected by nutrients when developing numeric nutrient criteria and secondarily focus on response variables that are indirectly affected by nutrients, such as primary production, and stream metabolism and respiration.

DWQ has already collected much algal and macroinvertebrate data. This includes the Utah State University Western Center for Monitoring and Assessment of Freshwater Ecosystems laboratory database (www.cnr.useu.edu). Most of these biological data can be indirectly used in the development and refinement of site-specific nutrient criteria metrics. However, relatively few taxa will occur at specific sites compared to the region-wide taxa pool and it is also well known that species, even within the same genus, can respond to environmental stressors in markedly different ways (e.g., Downes 2010, Macan 1963); therefore, generalized and regionalized taxonomies are not very useful and are poor guides to site-specific ecology. A thorough knowledge of the life histories and ecologies of these few taxa, particularly indicator taxa, is necessary and easily obtained from the literature.

Many metrics that are already in use in other assessment programs can be modified and, along with newly suggested metrics, can be combined for site-specific assessments. These metrics can include metrics based on species traits (e.g., functional feeding groups, life histories, voltinism, habitat associations, etc. (see Statzner and Beche 2010 and Vieira et al. 2006). Bioassessment programs are beginning to reconsider and integrate the ways in which environmental stressors can affect these species traits. Examples of potential metrics include:

Algal taxa richness and effective number of taxa

Macroinvertebrate taxa richness and effective number of taxa

Algal and macroinvertebrate assemblage structure

Total algal (periphyton) biomass

Total macroinvertebrate biomass (abundance corrected)

Individual indicator algal taxon biomass

Ex. *Cladophora* sp. and *Didymosphenia geminata*

Individual indicator macroinvertebrate biomass

e.g., scraper taxa or individual snail taxa

ex. *Potamopyrgus antipodarum* and *Corbicula* sp.

Functional feeding group ratios (both taxa and biomass based)

Secondary production rates of key indicator species of different functional feeding groups

Cold-water fish indicator species secondary production (e.g. growth rates)

Ex. *Cottus* sp., *Oncorhyncus* sp.

Changes in biomass, density and production are much better indicators of site-specific responses to nutrients than richness and abundance measures.

Macroinvertebrate Collections

Site-specific investigations should consider whether regional biological collection methods should be modified to better account for important site-specific characteristics. It has long been acknowledged (e.g., Barbour et al. 1999) that different floral and faunal assemblages respond over different spatial-temporal scales, and therefore respond to different scales of covariates. However, the framework of bioassessments has not yet fully utilized this realization. This is important because large amounts of variation in the structure of benthic communities can rarely be fully accounted for by regional relationships. On a site-specific basis several smaller-scale variables can cause dramatic shifts in community composition within a 1 m span of riffle (e.g., Brown 1972, Townsend 1989, Hart et al. 2013, Thomson et al. 2015). Although it is often assumed that large composite samples account for this kind of variation by "homogenizing the assemblage," this is not always the case (Marshall 2008), and site-specific study designs should carefully consider the pros and cons of composite samples.

The patch dynamics of species within a single riffle are affected by many variables such as the accumulations of fine and coarse detritus, the distribution and relative abundance of fine and coarse substrata, algal patch dynamics, and very local dynamics of flow (e.g., Barton and Smith 1984, Minshall 1984, Newbury 1984, Townsend 1989, Hansen et al. 1991, Hart and Finelli 1999, Lancaster and Downes 2010, Hart et al 2013). The relative importance of these sources of variation, among others, will depend on site-specific circumstances and the specific ecological response. The important point with respect to study designs is that local-scale habitat characteristics should all be considered when establishing site-specific collection methods for any of the ecological responses. For instance, near-substrata water velocity is known to have pervasive effects on benthic community structure (e.g., Hansen et al. 1991, Hart and Finelli 1999, Lancaster and Downes 2010 Thomson et al. 2015), which could be accounted for with a flow-stratified sampling regimen to prevent statistical confounding of the variable "velocity" with treatment effects.

Accounting for local-scale covariables along with species traits and other response variables (i.e., stream metabolism, water chemistry, and habitat assessments) should make algal and invertebrate assessment results more congruent with other lines of evidence. Reducing the uncertainty of responses associated with covariables should make all aspects of the assessment more informative by strengthening our understanding of the causal connections between nutrients or important covariables and these responses.

Temporal Variation

Spatial variation is often combined or confounded with temporal variation at several time scales however, there are several ways in which temporal variation should be accounted for (or controlled) in site assessments. Climatic variation is represented by changes among years in terms of temperature, rainfall, and other atmospheric characteristics (cloud cover, atmospheric pressure, storm frequency, etc.). It is best dealt with by sampling over a timeframe of several years. A BACI or a before-after controlled-impact paired series design (e.g., Osenberg et al. 1994) is useful because the differential responses of the study site and the internal reference site ("Control") can be used to estimate the degree of impact in all but the most extreme situations.

Seasonal variation can also complicate the interpretation of site assessments. For short-lived organisms (e.g., algae, bacteria, some invertebrates), frequent sampling is required to attain a reasonable estimate of the median annual condition. Frequent sampling helps to reduce the risk of uncertainty by preventing ephemeral shifts in production or taxonomic composition from being mistaken for chronic or acute problems. For longer-lived organisms (e.g., larger- sized macroinvertebrates or fish), the effects of seasonal variation are observed through the life cycles and life histories of animals, through emergence, or migrations. Isolating the effects of nutrients is difficult when a group of species are present one year and absent the next because of temperature-sensitive life-histories (e.g., Sweeney and Vannote 1984). The most common ways of reducing seasonal variation with longer-lived organisms is to sample during the same time period(s) each year. For macroinvertebrates, it is best to avoid time periods when many species are emerging from the system (spring). Thus, winter months are ideal—except for the interference of snow and ice. Although some insect species emerge in autumn, most of these species have multivoltine, asynchronous larval development and will not affect the richness or diversity as severely as the large emergences of univoltine insects occurring shortly before and shortly after spring runoff. These too can be accounted for through adequate sampling design (Osenberg et al. 1994).

Patch dynamics (Townsend 1989, Winemiller et al. 2010) are not only an important source of spatial variation in site assessments, they are also a source of temporal variation (seasonal and annual). For example, the structure and function of macroinvertebrate assemblages are intimately tied to flow (e.g., Hart and Finelli 1999), but the ideal refugia for a species may change over time. This can occur because of ontogenic shifts in the species' life history or simply because of natural physical changes in streams (e.g., spring discharge vs. late summer discharge). Thus, sampling at the exact same geographical location from one sampling period to the next may introduce unwanted variability because the habitat has functionally changed since the previous sampling period. The flow-stratified-systematic-sampling regimen alluded to elsewhere in this chapter (i.e., analysis of covariance; Milliken and Johnson 2002) circumvents this source of variation. Models based on regional data do not control for these types of temporal variation, leaving managers to assume all variance is due to stochastic processes or ecological impairments.

There is also an aspect of seasonal relevance for criteria. For example, if the purpose of a criterion is to protect larval fish, that criterion should be relevant during the period at which larval fish may be present.

Development of Detailed Sample Analysis Plans

Once indicators and sites have been selected a detailed sample analysis plan (SAP) can be created that will best reduce uncertainty. The SAP will define data quality objectives that take into consideration the many sources of variability discussed in this chapter. Specific collection methods that best account for important sources of variation will need to be described. Where possible, the analytical methods that will be used to derive goals should also be identified *a priori* because this will help determine the scope, sample frequency, and collection efforts that may be required. The SAP should also identify a process that allows the SAP to evolve. Finally, the SAP should specifically identify the roles and responsibilities for all parties involved in the investigation, including PIs, laboratory personnel, and field personnel.

Conclusion

Given the complexity of ecological systems, all sources of uncertainty in site-specific NNC development cannot be removed; therefore, resource managers need to decide on the degree of uncertainty that is acceptable in the context of specific management objectives. It is important to acknowledge that uncertainty that is not addressed often leads to conservative or even erroneous regulatory actions. As a result, it is in the interest of both DWQ and the regulated community to ensure that uncertainty is reduced to the greatest extent possible.

Chapter 16

USING QUAL2K MODELING TO SUPPORT NUTRIENT CRITERIA DEVELOPMENT AND WASTELOAD ANALYSES IN UTAH

B.T. Neilson,[1] A.J. Hobson,[1] N. vonStackelberg,[2] M. Shupryt,[2] and J. Ostermiller[2]

1. Utah State University, Civil and Environmental Engineering, Utah Water Research Laboratory,

2. Utah Department of Environmental Quality, Division of Water Quality

Key Points

Mechanistic models are important tools in the development of site-specific numeric criteria.

Qual2Kw is a quasi-dynamic, one-dimensional instream water quality model that is often used for regulatory purposes because it includes the dominant processes of concernwhen evaluating the effects of nutrient enrichment.

The large number of processes included in the model results in a large number of variables that can complicate model calibration.

A sensitivity analysis was conducted to evaluate the relative importance of model variables to help prioritize future data collection efforts.

It is recommended that model calibration data are collected over the longest practicable time period.

Empirical calibration, as opposed to autocalibration, of key parameters such as biological oxygen demand, nitrification rates and water column photosynthetic active radiation would also improve the accuracy of model interpretation.

Foreword

Division of Water Quality (DWQ) funded an investigation to identify how Qual2K mechanistic models could be used to help establish site-specific numeric nutrient criteria (NNC), or to confirm the applicability of regionally-derived NNC on a site-specific basis. This report provides specifications for application of the results of these models to NNC and related nutrient reduction regulatory programs. Implementation of these specifications will differ on a case-by-case basis and DWQ will provide specific information about any application of these models as part of the documentation that accompanies any regulatory actions that are taken based on the specifications in this report.

Background

Over the last few years the State of Utah has been working toward understanding the implications of instituting NNC. To do this, the state initiated a publicly owned treatment works (POTW) nutrient-removal cost study (Utah Department of Environmental Quality [UDEQ] 2009), an economic evaluation study (UDEQ 2011), and a nutrient criteria ecological study (UDEQ 2010). The *Nutrient Removal Cost Impact Study*, completed in 2009, evaluated the economic impacts of potential new nutrient-removal requirements for Utah's POTWs. The study estimated economic, financial, and environmental impacts interrelated with a range of potential nutrient discharge standards for every discharging mechanical POTW in the state and one lagoon system (UDEQ 2009). The economic evaluation study, is intended to quantify the economic benefits and costs of implementing nutrient criteria for surface waters in Utah (UDEQ 2011).

When investigating nutrient criteria based on ecological implications, the U.S. Environmental Protection Agency (USEPA) recommends three types of scientifically defensible empirical approaches for establishing numeric thresholds intended to limit nitrogen (N) and phosphorus (P) pollution: reference condition approaches, mechanistic modeling, and stressor-response (S-R) analysis (USEPA 2010). UDEQ is currently investigating all three recommended approaches to establish NNC. In order to complete two of the recommended approaches, the UDEQ, along with Utah State University (USU), is investigating the ecological impacts of nutrients on Utah waterbodies using an S-R approach combined with the predictive capabilities of the mechanistic modeling approach. To do this, a data collection strategy was developed to meet the needs of both recommended approaches. By combining the results of the economic study with the predictive capabilities of the modeling efforts and ecological response information, the proposed instream nutrient criteria can be linked to the expected economic costs of the treatment upgrades and forecast the potential impact of nutrient loading on the ecological health of the downstream waterbodies.

This report covers the general approaches taken in the mechanistic modeling portion of the nutrient criteria study for data collection, model population, and calibration/validation. It is important to note that the data collection approaches and models are intended to have multiple applications and, therefore, have been made very generic in order to support: (1) development of statewide NNC, (2) development of site-specific criteria for rivers and streams where the statewide nutrient criteria do not appear valid, (3) wasteload analyses to determine water quality based effluent limits (WQBEL), and (4) determination of total maximum daily load (TMDL) endpoints. Details regarding the associated ecological measures and reference condition information can be found at http://www.nutrients.utah.gov/nutrient/index.htm.

QUAL2Kw Model

UDEQ uses low-flow conditions to determine WQBELs for point sources under the Utah Pollution Discharge Elimination System program (UDEQ 2012), partly because these low-flow conditions correspond with limiting conditions. This led to selection of a model that would be appropriate for these conditions. QUAL2K (Chapra et al. 2004) is an USEPA-approved model that has been commonly used in wasteload allocations (WLAs) and TMDLs (e.g., Bischoff et al. 2010, Kardouni and Cristea 2006) and in development of nutrient criteria (Flynn and Suplee 2011). This quasi-dynamic, one-dimensional instream water quality model includes the dominant processes of concern for Utah waters, predicts the required water quality variables, and is feasible to populate and calibrate given the limited data available in most waterbodies of the state. To understand the associated daily minimum and maximum instream concentrations, the model provides a 24-hour diel response in water quality given an appropriate or representative 24-hour weather pattern. QUAL2Kw (Pelletier and Chapra 2008), a sister model to QUAL2K developed within the State of Washington, built in additional functionality (e.g., automatic calibration algorithms) based on their identified needs. With the expectation of having some needs similar to Washington and the possibility of identifying additional needs, UDEQ elected to use QUAL2Kw in their instream modeling applications.

Details regarding the version of QUAL2Kw used in this application (version 5.1) are provided within the user's manual (Pelletier and Chapra 2008) and a number of publications (Cho and Ha 2010, Kannel et al. 2007, Pelletier et al. 2006). In short, the state variables (Table 16.1) include the macro nutrients of interest (carbon [C], N, and P) and the critical nutrient species (e.g., inorganic P, nitrate, and ammonia) in surface waters.

Table 16.1. QUAL2Kw state variables (taken directly from Pelletier and Chapra 2008).

Variable	Symbol	Units*
Conductivity	s_1, s_2	μmhos
Inorganic suspended solids	$mi_{,1}$, $mi_{,2}$	mgD/L
Dissolved oxygen	o_1, o_2	mgO_2/L
Slow-reacting carbonaceous biochemical oxygen demand (CBOD)	$cs_{,1}$, $cs_{,2}$	mg O_2/L
Fast-reacting CBOD	$cf_{,1}$, $cf_{,2}$	mg O_2/L
Organic nitrogen	$no_{,1}$, $no_{,2}$	μgN/L
Ammonia nitrogen	$na_{,1}$, $na_{,2}$	μgN/L
Nitrate nitrogen	$nn_{,1}$, $nn_{,2}$	μgN/L
Organic phosphorus	$po_{,1}$, $po_{,2}$	μgP/L
Inorganic phosphorus	$pi_{,1}$, $pi_{,2}$	μgP/L
Phytoplankton	$ap_{,1}$, $ap_{,2}$	μgA/L
Detritus	$mo_{,1}$, $mo_{,2}$	mgD/L
Pathogen	x_1, x_2	cfu/100 mL
Generic constituent	gen_1, gen_2	user defined
Alkalinity	Alk_1, Alk_2	$mgCaCO_3/L$

Variable	Symbol	Units*
Total inorganic carbon	$cT_{,1}$, $cT_{,2}$	moles/L
Bottom algae (a_b in the surface water layer), biofilm of attached heterotrophic bacteria (a_h in the hyporheic sediment zone for the Level 2 option)	a_b,a_h	gD/m^2
Bottom algae nitrogen	IN_b	mgN/m^2
Bottom algae phosphorus	IP_b	mgP/m^2

* mg/L ≡ g/m^3

Using the same notation as that of Table 16.1, the QUAL2Kw composite or calculated variables are (Pelletier and Chapra 2008):

Total Organic Carbon (mgC/L):

$$TOC = \frac{c_s + c_f}{r_{oc}} + r_{ca}a_p + r_{cd}m_o \quad (1)$$

Total Nitrogen (μgN/L):

$$TN = n_o + n_a + n_n + r_{na}a_p \quad (2)$$

Total Phosphorus (μgP/L):

$$TP = p_o + p_i + r_{pa}a_p \quad (3)$$

Total Kjeldahl Nitrogen (μgN/L):

$$TKN = n_o + n_a + r_{na}a_p \quad (4)$$

Total Suspended Solids (mgD/L):

$$TSS = r_{da}a_p + m_o + m_i \quad (5)$$

Ultimate Carbonaceous BOD (mgO$_2$/L):

$$CBOD_u = c_s + c_f + r_{oc}r_{ca}a_p + r_{oc}r_{cd}m_o \quad (6)$$

Additionally, the model provides the ability to predict the associated biological effects of various nutrient concentrations since photosynthesis, respiration, and death of phytoplankton and bottom algae are included within the model. As the version of QUAL2Kw applied within this study is quasi-dynamic, it provides the ability to deal with steady flow while allowing for nonuniform flow. This means that while the flow conditions cannot change over time, they can vary longitudinally downstream due to point or distributed inflows or abstractions.

Given the capabilities of this version of QUAL2Kw, the approach is suited to particular environmental conditions. The time periods over which this model should be applied require that (1) stream conditions are completely mixed since the model assumes all model elements are completely mixed, (2) boundary condition concentrations can be approximated by consistent 24-hourly values, (3) distributed flows are constant, (4) point inflows follow a consistent diel pattern or are constant, and (5) weather conditions over the simulation period have a consistent diel pattern.

Study Site Locations

Nine sites were selected for the nutrient criteria ecological study and represent the different types of receiving waterbodies around the State of Utah (Figure 16.1). Using these sites as a representative

Figure 16.1. Study site locations within the State of Utah.

sample of the state's waterbodies, the QUAL2Kw, ecological S-R, and reference condition findings will be used to extrapolate information regarding possible ranges of nutrient criteria for the remaining state waters (UDEQ 2010). The selected sites (Table 16.2) are located within different-order streams that vary in background water quality, in their surrounding land uses, in the amounts of wastewater effluent, and in the levels to which the effluent has been treated. The sections studied were those influenced by water reclamation facility (WRF) effluents since these areas generally have enhanced nutrient loads.

Table 16.2. Study site locations, water reclamation facilities, and dates sampled within the State of Utah.

Waterbody	Facility	Dates Sampled
Box Elder Creek	Brigham City WRF	Aug. 9–11, 2010
San Pitch River	Fairview City WRF	Aug. 2–5, 2010 Oct. 11–13, 2010
San Pitch River	Moroni City WRF	July 28–30, 2010
Weber River	Oakley City WRF	Aug. 23–26, 2010
Price River	Price River Water Improvement District	Aug. 30–Sep 1, 2010
Dry Creek	Spanish Fork City WRF	July 23–26, 2010
Silver Creek	Snyderville Basin-Silver Creek Water Reclamation Facility	July 20–22, 2010 Sep. 30–Oct. 4, 2010 Aug. 22–30, 2011
Malad River	Tremonton City WRF	Aug. 13–16, 2010
Little Bear River	Wellsville Lagoons	Sept. 10–13, 2010

More details about each site (e.g., location, study reach length, etc.) are provided in a separate report in preparation by DWQ that evaluates structural and functional responses to nutrients. Detailed information about unique sampling requirements associated with each site and the specific information regarding model population, calibration, and validation are provided within the QUAL2K modeling files and site-specific model documentation provided to UDEQ as project deliverables.

Project Results

Since the key objectives in this project were to develop the appropriate data collection methodologies to support QUAL2Kw model population and calibration, this report provides general information regarding the field data requirements, approaches to model population using these data, strategies used in model calibration, and the steps required for model validation (for cases where these datasets exist).

Supporting Field Data

Data must be collected at three types of locations for instream modeling: the beginning of the study reach (also called the headwaters or upstream boundary condition), inflows/outflows (point sources or tributary inflows and diversions/abstractions), and at least one location downstream for model calibration. The data types required at these locations will vary and are discussed below. For the 2010 data collection efforts, data collection at each location spanned a 2-day period during low-flow conditions.

Figure 16.2 shows a generalized schematic used with the 2010 data collection efforts. Data for the modeling efforts were gathered at Station B (headwaters/upstream boundary condition), Station C (wastewater treatment plant effluent before it enters the stream), Station D (a location where the stream and point source effluent were completely mixed), and Station E (the calibration location downstream at the end of the study reach). Information gathered at Station A was only used in the ecological portion of the study. If a tributary entered the modeling reach, data were also collected at T1. Similarly, if a diversion was present, the quantity of water leaving the system was determined at D1.

For the 2010 data collection, the location of the completely mixed conditions downstream of the WRF was determined by measuring specific conductance or temperature across the channel to determine where uniform conditions existed. In some cases where the differences in temperature and/or specific conductance were too small, rhodamine water tracing (WT) was used as a visual indicator. To support and integrate these efforts with the ecological study needs, the distance between Station D and E was estimated using methods described in Grace and Imberger (2006), which designate the optimum distance between stations for calculating open water metabolism using the single station method (Equation 7).

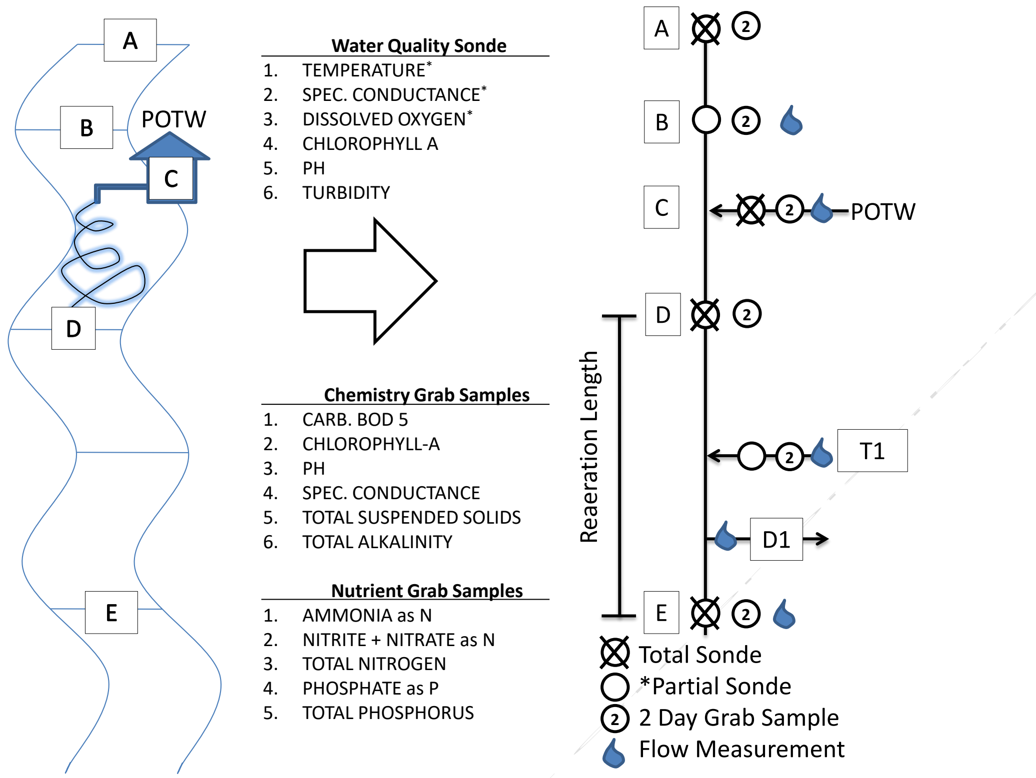


Figure 16.2. Generalized data collection locations for the 2010 sampling efforts. Required locations of flow measurements and the multi-parameter water quality sondes are also shown. Since 2010 modeling efforts used Station B as the headwater location, the information for chlorophyll-*a* and pH is taken from Station A.

$$X = 0.693 \cdot \frac{v}{k_a} = \frac{v^{0.33}}{0.0137 \cdot \left(D^{-0.85}\right)} \quad (7)$$

Where, X = optimum station distance (km), v = velocity (cm s^{-1}), D = depth (cm), and k_a = reaeration coefficient for oxygen (d^{-1}).

As discussed later, this method was found to result in distances that were generally too short to meet the diverse needs of this study and often did not include the compliance point for WLAs.

The information that needs to be collected at each of these stations depends on whether it is the headwater location, a load, or a diversion. Water quality models require an understanding of both the water balance and the mass balances for each constituent modeled. Flow measurements may be required at all stations to establish a water balance. Water quality information is not, however, required for diversions or abstractions since the mass loss will be a function of the instream concentrations predicted by the model and the volume of water taken out that is specified by the user.

The specific water quality constituents measured at each station and the frequency at which they were collected are detailed in Table 16.3. A number of constituents (e.g., temperature, dissolved oxygen [DO], conductivity) were measured using multi-parameter sondes at 5-minute increments over each of the two-day sampling periods. Grab samples of most constituents requiring laboratory analyses were gathered each day. usually in the mid-morning. Benthic algae sampling was only conducted once at some point in time close to the study periods. A number of constituents within this list, indicated by an asterisk in the table, were not sampled directly and had to be estimated. The appropriate values for modeling were estimated using the relationships between measured constituents and model variables as described below. Additional data types that could be collected that would be useful in the modeling include a measure of sediment oxygen demand, total organic carbon, and volatile suspended solids. None of these data types were collected in the 2010 datasets.

Table 16.3. Water quality constituents sampled and the frequency of sampling for QUAL2Kw modeling.

Multi-Parameter Sonde Data	Abbreviation/QUAL2Kw Units	Frequency
Field Perameters		
Water temperature	Temp (C)	5-min samples
Specific conductance	COND (mhos)	5-min samples
Dissolved oxygen	DO (mgO_2/L)	5-min samples
pH	pH	5-min samples
Chlorophyll-*a*	CHLA (gA/L)	5-min samples
Turbidity		5-min samples
Laboratory Analysis		
5-Day soluble carbonaceous BOD, $sCBOD_5$		1 each day
Total nitrogen	TN (gN/L)	1 each day
Ammonia nitrogen	NH_4 (gN/L)	1 each day
Nitrate + nitrite nitrogen	NO_3 (gN/L)	1 each day
Total phosphorus	TP (gP/L)	1 each day
Soluble reactive phosphorus	SRP (gP/L)	1 each day
Volatile suspended solids*	VSS (mgD/L)	1 each day
Total suspended solids	TSS (mgD/L)	1 each day
Alkalinity	ALK ($mgCaCO_3$/L)	1 each day
Chlorophyll-*a*	CHLA (gA/L)	1 each day
Dissolved organic carbon (DOC)	DOC (mgC/L)	1 each day
Dissolved organic phosphorus (DOP)*		1 each day
Dissolved organic nitrogen (DON)*		1 each day
Benthic Chlorphyll-*a*		1 per sampling time period
Benthic ash free dry mass (AFDM)		1 per sampling time period
Benthic total phosphorus (TP)		1 per sampling time period
Benthic total nitrogen (TN)		1 per sampling time period
Benthic total organic carbon (TOC)		1 per sampling time period

Multi-Parameter Sonde Data	Abbreviation/QUAL2Kw Units	Frequency
Sediment Oxygen Demand (SOD)#		1 per sampling time period
Total Organic Carbon (TOC)#	TOC (mgC/L)	1 each day

* = not gathered or required estimation for QUAL2Kw
# = data that would be useful in model population/calibration but were not directly measured in these efforts

Data to characterize each site are necessary to support model population or calibration. Table 16.4 provides a list of the data types requiring collection, some procedural information, locations where these data are required within or near the site, and the utility of the data in the context of the modeling effort. A number of these data types are collected during routine surveys for Utah's Comprehensive Assessment of Stream Ecosystems (UCASE) fieldwork, based on protocols adapted from USEPA (2007). It is important to note that GPS coordinates must be established at all locations where data are collected, to ensure comprehensive documentation of the data collection process.

Table 16.4. Site characterization data types.

Data Type	Procedure	Locations	Reasoning
Average Cross-Sectional Velocity*	See methods provided with data collection and/or UCASE standard operation procedures (SOP). Information from HEC-RAS modeling applications can also be extracted to supplement data collected.	Stations D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of velocity in different reaches to compare with the predicted velocities. This can be used with the depth and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Cross-Sectional Depth*	See methods provided with data collection and/or UCASE SOP.	Stations D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of depths in different reaches to compare with the predicted depths. This can be used with the velocity and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Channel Bottom Width	Bottom width estimates were calculated using side slope, average depth, and top width values in the formula: Top Width – Depth x 1/tan(radians(°SSLEW)) – Depth x 1/tan(radians (°SSREW)), where width and depth are in meters and side slope is in radians in the form of Run/Rise.	Stations D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input.
Channel Bottom Slope	See methods provided with UCASE SOP.	Should estimate bottom slope from beginning to end of study reach at 10% increments of total reach length and/or when changes in bottom slope are observed.	Model input.
Channel Side Slope	See methods provided with UCASE SOP.	Stations D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input can be used to calculate bottom width from measured top widths.

Data Type	Procedure	Locations	Reasoning
Weather Data	Onsite weather station or nearest Mesowest Station.	Near study site would be most appropriate and 15–30-minute data are preferred.	Wind speed, air temperature, shortwave solar radiation, and humidity/dewpoint temperature are all used within the model as forcing information. Precipitation data show whether there was significant rainfall in the area that would influence instream flows.
Tracer Study	Inject tracer at Stations B or C and measure response at Station E. Can also use HEC-RAS model if available.	Measure tracer response at Station E, but additional locations along the study reach would be beneficial to capture heterogeneity.	Provides information regarding average travel time through system and can be used in calibration of hydraulic parameters (e.g., Manning's roughness coefficient).
Substrate Type*	See methods provided within Data Collection and/or UCASE SOP.	Information should be gathered at cross sections in subreaches that represent the variability in substrate types.	Provides a method to approximate the Manning's roughness coefficient and determine fraction of bottom substrate appropriate for bottom algae.
Shading*	See methods provided with data collection and/or UCASE SOP.	Information should be gathered at locations that represent the variability in shading.	Model input. If riparian or topographic shading drastically influences instream temperatures, estimates of the shading % for each hour of a day will be necessary to scale the incoming shortwave solar radiation.

* = not gathered or required estimation for QUAL2Kw

Model Population

Once the data have been collected, they must be translated from observations to model inputs. The model state variables (Table 16.1) can be related to measurements are as follows (taken directly from Pelletier and Chapra [2008]):

$$\text{Conductivity} = s = \text{COND} \quad \textbf{(8)}$$
$$\text{ISS} = m_i = \text{TSS} - \text{VSS or TSS} - r_{dc}\,(\text{TOC} - \text{DOC}) \quad \textbf{(9)}$$
$$\text{Dissolved Oxygen} = o = \text{DO} \quad \textbf{(10)}$$
$$\text{Organic Nitrogen} = n_o = \text{TKN} - \text{NH4} - r_{na}\,\text{CHLA} \quad \text{or} \quad \textbf{(11)}$$
$$n_o = \text{TN} - \text{NO2} - \text{NO3} - \text{NH4} - r_{na}\,\text{CHLA}$$
$$\text{Ammonia Nitrogen} = n_a = \text{NH4} \quad \textbf{(12)}$$
$$\text{Nitrate Nitrogen} = n_n = \text{NO2} + \text{NO3} \quad \textbf{(13)}$$
$$\text{Organic Phosphorus} = p_o = \text{TP} - \text{SRP} - r_{pa}\,\text{CHLA} \quad \textbf{(14)}$$
$$\text{Inorganic Phosphorus} = p_i = \text{SRP} \quad \textbf{(15)}$$
$$\text{Phytoplankton} = a_p = \text{CHLA} \quad \textbf{(16)}$$
$$\text{Detritus} = m_o = \text{VSS} - r_{da}\,\text{CHLA or } r_{dc}\,(\text{TOC} - \text{DOC}) - r_{da}\,\text{CHLA} \quad \textbf{(17)}$$
$$\text{pH} = \text{PH} \quad \textbf{(18)}$$
$$\text{Alkalinity} = Alk = \text{ALK} \quad \textbf{(19)}$$

While a number of these relationships are straightforward, it is important to realize that the typical Organic N and Organic P measurements cannot be directly compared to the Organic N and Organic P QUAL2Kw predictions. As shown in equations 11 and 14 above, the QUAL2Kw versions of Organic N and P only represent the dissolved and detritus portion of each organic nutrient pool since the portions associated with the live algae are subtracted out. It is also important to note that detritus (Equation 17) only contributes to the carbon budget and does not influence other nutrient pools.

Given the data available from the 2010 sampling, additional methods were needed based on some assumptions or established equations to calculate the variables necessary for model population and calibration. These included the need to convert $sCBOD_5$ measurements to the sCBOD values required within QUAL2Kw (Equation 20).

$$\text{sCBOD ultimate} = c_f \text{ or } c_s = \text{sCBOD}_5/(1-\exp(-k_d\ (5\text{days})) \qquad \textbf{(20)}$$

Since direct measures of VSS or inorganic suspended solids (ISS) were not collected for the 2010 datasets, methods and a logic tree were developed to estimate these values for the model population

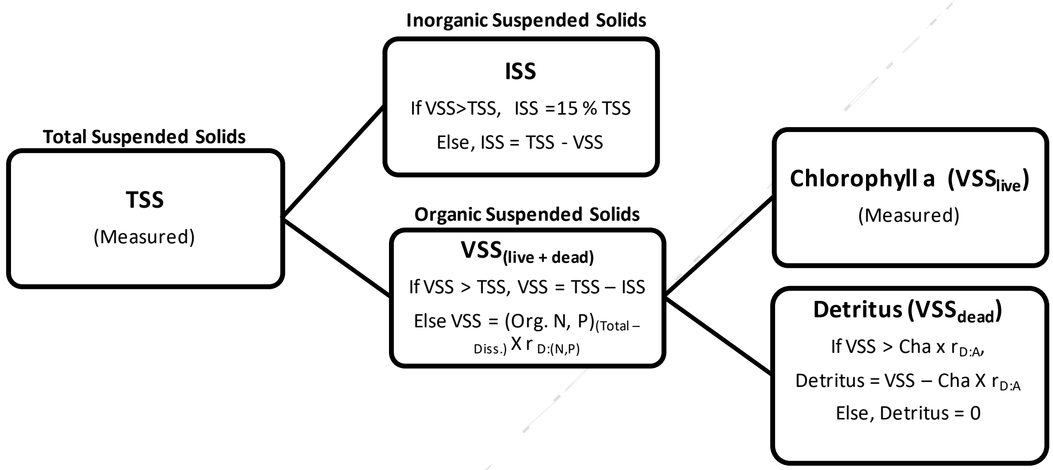


Figure 16.3. Logic used in estimating volatile suspended solids (VSS) and inorganic suspended solids (ISS) from total suspended solids (TSS), followed by logic for estimating detritus from VSS.

(Figure 16.3).

To populate the model, information regarding the reach, initial conditions, headwater conditions, weather data, point sources, and distributed sources must be provided (Table 16.5). More specifically, observations from the headwater location and any point flow (inflow or abstraction) or distributed flows must be entered into the model framework. Any flow information provided for these locations must be a representative value for the entire modeling period. The necessary sampling frequency of specific water quality data is dependent upon whether it is a point source or headwaters (Table 16.6). The other forcing data type required by the model is meteorological information, which includes hourly average air temperatures, wind speeds, and dewpoint temperatures from a nearby representative weather station. Shortwave solar radiation can be estimated automatically within the modeling framework; however, if using these estimates, hourly cloud cover values would be required. In the 2010 modeling efforts, actual shortwave radiation observations from a local source were used.

Table 16.5. General information required for QUAL2Kw model population.

QUAL2Kw Sheet	Information Required
Reach	Reach segmentation Hydraulic characteristics % suitable substrate Bottom algae % cover Sediment oxygen demand (SOD) Thermal properties
Initial Conditions	Constituent concentrations (See Table 16.6)
Headwater Data	Average flow Constituent concentrations (See Table 16.6)
Weather Data (hourly average values)	Air temperature Dewpoint temperature Solar radiation Shading Cloud cover Wind speed
Point Sources	Average flow Constituent concentrations (See Table 16.6)
Distributed Sources	Average flow Constituent concentrations (See Table 16.6)
Rates	Primarily set in calibration. See "Model Calibration" section below.

Table 16.6. Model input constituent concentration requirements and the associated observed data used in population of QUAL2Kw.

Model Parameter	Data Collected	Point Source	Headwater	Distributed Inflow
		Mean + Range/2 or 2-Day Mean	Hourly Average or 2-Day Mean	Average
Alkalinity	Total Alkalinity	X	X	X
$sCBOD_{ultimate}$	$sCBOD_5$	X	X	X
Specific Conductivity	Specific Conductivity	X	X	X
Detritus (Particulate Organic Matter [POM})	(Org - Diss. N, P) X r(POM/N, P)	X	X	X
Dissolved Oxygen	Dissolved Oxygen	X	X	X
Inorganic Phosphorus (SRP)	Inorganic Phosphorus (SRP)	X	X	X
Inorganic Solids	TSS - VSS	X	X	X
NH_4 - Nitrogen	NH_4 - Nitrogen	X	X	X
NO_3 - Nitrogen	NO_3 - Nitrogen	X	X	X
Organic Nitrogen	TN - (NH_4) - (NO_3 + NO_2)	X	X	X
Organic Phosphorus	TP - Inorg P	X	X	X
pH	pH	X	X	X
Phytoplankton	Chlorophyll a	X	X	X
Water Temperature	Water Temperature	X	X	X

When populating the models, censored data, or concentrations that are below the analytical detection limits (i.e., nondetects), commonly occur. Within the 2010 modeling study, nondetects were assigned a concentration of one-half of the detection limit. More accurate statistical analysis of limited amounts of censored data should be investigated. The detection limits associated with key parameters are detailed within Table 16.7.

Table 16.7. Detection limits for constituents based on the procedures applied within specific laboratories.

Constituent	Laboratory	Analytical Detection Limit (mg/L)
TN, TP	Baker Lab—USU	0.0057
TDN, TDP	Baker Lab—USU	0.0025
$NO_3 + NO_2$ - N	Baker Lab—USU	0.0006
NH_4 - N	Baker Lab—USU	0.00395
PO_4 - P	Baker Lab—USU	0.0008
$sCBOD_5$	UDEQ Laboratory/AWAL	3/5
Chlorophyll-*a*	UDEQ Laboratory	0.0007
Specific Conductance	UDEQ Laboratory	2 (uS/cm)
Total Suspended Solids	UDEQ Laboratory	4
Total Dissolved Solids (180°C)	UDEQ Laboratory	10
Turbidity	UDEQ Laboratory	0.1 (NTU)

Note: TDN = total dissolved nitrogen, TDP = total dissolve phosphorus, TN = total nitrogen, TP = total phosphorus

To assist in ensuring model population consistency given the relatively consistent data collection strategies implemented in 2010, two supporting sheets within the QUAL2Kw files were developed and delivered to UDEQ. A "Data Input" and "Addt Info" sheet provide a number of tables that can be populated with observations, and this information automatically populates the QUAL2Kw sheets. These sheets facilitate some of the additional calculations that were completed and suggested for future applications (described further below).

Most information within the Rates Sheet was not changed at all or was adjusted in model calibration (described further below). However, specific values of some parameters were established within the 2010 modeling efforts that may be appropriate for other Utah model applications. First, CBOD decomposition rates (k_d) were measured by taking 6 samples from the Silver Creek WRF effluent. These samples were analyzed in triplicate, resulting in 18 total measurements of 30-day CBOD using methods detailed in a report by the Environmental Protection Division (1989). The resulting data were analyzed using a nonlinear least squares method and the Thomas method (Table 16.8). Given that Chapra (1997) reports values ranging from 0.05 to 0.1 d^{-1} at 20°C for waste streams treated using activated sludge, the average value of 0.103 d^{-1} was assumed to be an appropriate value for all the 2010 study sites; therefore, this value was not varied in calibration. This value was used to convert any measured concentrations of $sCBOD_5$ to the sCBOD values required by the model (Table 16.6). A Utah-specific number should be established for the dominant wastewater treatment types of activated sludge and membrane mechanical treatment, as well as for lagoon systems.

Table 16.8. Carbonaceous biochemical oxygen demand decomposition rate statistics based on samples from Silver Creek WRF effluent.

Statistic	NLS Method *kd*, 1/d	Thomas Method *kd*, 1/d
Min	0.095	0.076
Max	0.125	0.124
Mean	0.103	0.096
StDev	0.011	0.012
95% CI	0.013	0.013

The thermal properties of the Silver Creek substrate were also measured since these dictate the rate of heat exchange between the water column and the sediments. While a number of values are reported within the QUAL2K model software and QUAL2Kw manual, it can be important to have site-specific thermal properties. The thermal property values based on measurements from Silver Creek with a sandy-gravel substrate were a thermal diffusivity of 0.72 $mm^2 s^{-1}$ and a thermal conductivity of 2.25 $W m^{-1} k^{-1}$.

Model Calibration

Calibration within this effort consisted of a number of manual calibration steps followed by autocalibration using the genetic algorithm within QUAL2Kw. The data used in calibration included hydraulics data (longitudinal depths, velocities, and travel time) and water quality data (Table 16. 3) including the mean, minimum, and maximum values at each calibration location (only station E for the 2010 effort). For those data types where only two samples were taken, the minimum and maximum values were not always representative of the daily variability and only provided an understanding of the range at these sampling times.

Manual Calibration Steps

A number of manual calibration steps or checks were used to ensure that the model was representing the system well based on site-specific data. These steps are key since they ensure that the foundational model components (e.g., flow balance, volumes, and instream temperatures) are correct before moving on to the more interconnected mechanisms associated with nutrient cycling.

Flow Balance and Hydraulics

Several steps were taken to ensure that the representation of the hydraulics was appropriate; many data types must be considered in this process. To ensure discharge matched empirical observations, predicted values were compared to measured values. If the values differed, the cause could have been groundwater exchanges or inflows or outflows from unknown sources. Although the reaches in these studies did not have significantly large differences, in some cases it may be necessary to incorporate a distributed inflow or abstraction to represent groundwater influences.

Specific conductance values were used in a number of different ways. Predicted values were compared to observed specific conductance values including the diel fluctuations. Since specific conductance is a relatively conservative measure of dissolved species, if predictions did not match the observations, this could indicate the presence of unknown inflows and may suggest the need for additional time in the field to determine the source of the inflow.

Travel times within the study reach are dependent on having the channel geometry, water depth, and velocities correct. After data collection efforts were completed, a tracer study was conducted using either salt or rhodamine WT to provide data regarding travel times within the study reaches. Because Manning's equation is used to route the water through the study reach, additional information must be provided at a subreach scale about bottom widths, side slope, channel bottom slope, and Manning's roughness coefficient. Top widths, side slopes, and bottom slopes are measured at consistent increments along the channel. From these data, as described in Table 16.4, bottom width estimates were calculated using side slope, average depth, and top width values. Once the bottom width, side slope, and bottom slope values were entered into the model and the model was run, the predicted top widths were compared to field-derived data. If necessary, the bottom widths or side slopes were adjusted, within reason. While good estimates of water depth and velocity were available at a number of discharge measurement locations, these values were not always recorded. After model setup, where available, predicted water depths and velocities were compared to measurements from locations downstream. At the same time, predicted travel times were compared to those estimated from tracer injection responses at various locations downstream. If necessary, Manning's n and possibly bottom slope were adjusted to ensure water depths, velocities, and travel time predictions were similar to observations. Once the hydraulic representation was appropriate, the temperature and ISS predictions were assessed to determine whether their accuracy was acceptable.

Temperature

First, predicted and observed temperatures at different locations downstream were compared. If predictions were inaccurate, shading data and trial and error approaches were used to adjust the hourly percent-shading values. Another consideration was the accuracy of the predicted top widths. This was key because the predicted surface heat flux values depend on the surface area of the air–water interface. At times, it may be necessary to revisit the top width predictions to ensure the accuracy of temperature predictions. Additionally, when inflows are present, the temperatures of the inflows may need to be adjusted if they were not measured in the field.

Inorganic Suspended Solids

Second, predicted and estimated ISS concentrations were compared at various locations longitudinally. The settling velocity was adjusted to vary the predicted concentrations. Ensuring these values were correct was important for accurately modeling photosynthesis rates due to the influence of ISS on light penetration. In this study, the ISS observations were calculated, not directly measured, and it is unclear whether they are accurate.

Reaeration Rates

Finally, to minimize the number of parameters that are varied in the autocalibration, an approach was developed to determine the appropriate reaeration formula to apply within the model, and a method was developed to approximate SOD using the DO timeseries collected at each site. To determine a representative reaeration formula, whole stream metabolism methods were applied to estimate gross primary production (GPP) and ecosystem respiration (ER) using the concentrations at most stations where DO was measured within the study reach. As part of this, it is necessary to estimate a reaeration rate (k_a). Various "open-water" methods of determining GPP, ER, and k_a have been established, including the delta method (McBride and Chapra 2005), night time regression (Young et al. 2004), and inverse method (Holtgrieve et al. 2010). It is possible to select the appropriate method for sites based on the recommendations of Aristegi and colleagues (2009) where the delta method was found to be best in open canopy and clear conditions (using the point method if data are smooth and the centroid method if data are noisy) and the night time method was more appropriate in turbulent reaches and where WRF effluent is dominant and there are highly variable flows. For various sites in Utah, the inverse method was found to produce the most consistent results based on estimates for many systems and sites across the state.

To support the QUAL2Kw modeling, the night time regression and inverse methods were applied to estimate the GPP, ER, and/or k_a at locations along the study reach. Given the variability of predicted reaeration rates from the formulas included within QUAL2Kw and the associated uncertainty, a number of steps were taken to determine the most appropriate formula. First, the QUAL2Kw model was run using each reaeration formula. The predicted reaeration rates in each reach segment were compared to the k_a values estimated from the metabolism methods, and a root mean square error (RMSE) was calculated. Next, the most appropriate formula based on the lowest RMSE value was calculated and set within the model. If multiple equations were appropriate, the one meeting all assumptions (e.g., depth and velocity ranges) was selected.

As described within Appendices A and B, these steps have been automated within the "Data Input" sheet that USU added for the 2010 modeling study. It is important to note that it would be possible to set the reaeration rates based on the values obtained from the metabolism measurements directly; however, because these estimates are derived empirically, it may limit the ability of the model to accurately predict reaeration under flow conditions that differ from those observed during metabolism data collection (i.e., due to different velocities and depths) and within reach segments where k_a values were not measured.

Sediment Oxygen Demand

Another significant source of uncertainty in QUAL2Kw modeling, particularly in shallow streams, is the amount of SOD present within each system. While QUAL2Kw has the functionality to estimate SOD based on a sediment diagenesis algorithm, there is often more SOD present than is predicted. The need to prescribe SOD has been associated with the deposition of organic matter outside of the time period of the model simulation (i.e., during snowmelt runoff), and the deposition of coarse particulate organic matter

(CPOM) that typically is not captured by standard sampling techniques. This extra SOD became an issue within the Jordan River TMDL (Stantec Consulting 2010) and was addressed through direct measurements of SOD to determine reasonable ranges that would be acceptable within QUAL2Kw modeling. In many cases, however, these types of measurements will not be available due to cost and the personnel requirements to collect them, and the results have significant variability. Change in oxygen over time is known to be a result of oxygen sources (primary production and reaeration) and oxygen sinks (autotrophic and heterotrophic respiration, BOD, and other oxygen-consuming reactions within the water column and sediments). However, when using metabolism methods, the equation describing the change in oxygen is reduced to:

$$dO/dt = GPP + reaeration - ER \quad \textbf{(21)}$$

Where,

GPP = gross primary production
ER = ecosystem respiration.

In this context, *ER* is now a net sink term. If autotrophic respiration is assumed to approximately equal *GPP* (it may need to be some fraction of *GPP* (Jones et al. 1997), then any extra oxygen consumption is due to heterotrophic respiration and other oxygen-consuming reactions within the sediments and water column. If this value is positive (meaning *ER* is higher than *GPP*), this provides an estimate of total SOD (heterotrophic respiration + oxygen-demanding reactions within the sediments) and some oxygen-demanding reactions within the water column (e.g., BOD decomposition and nitrification). Within QUAL2Kw, it can be assumed that this total SOD value would provide a maximum SOD that could be prescribed within the model. In most cases, the maximum SOD should include the prescribed SOD plus the SOD estimated within the sediment diagenesis algorithm within QUAL2Kw (described within Pelletier and Chapra [2008]). In these efforts, the *ER* minus *GPP* approximation for SOD was assumed to be appropriate since the streams included in this study are relatively shallow, and sediment processes will significantly influence the water column DO response. In larger rivers, it is possible that other processes more significantly influence the water column oxygen responses (e.g., chemical reactions within the water column, phytoplankton, etc.), and these approaches may not be applicable or may include more error due to the aforementioned assumptions.

Since SOD measurements were not gathered during the 2010 data collection efforts, *ER* values minus the *GPP* estimates were used at Station E (and at times at Station D) to determine a reasonable average and range of SOD values for the portion of the study reach below the WRF. Where appropriate, an average value of SOD was established and set within the model before autocalibration.

Autocalibration

With a number of parameters set based on the prior manual calibration steps, the remaining parameters that were appropriate to include in model calibration were autocalibrated. The parameters that should be included in calibration and the appropriate parameter ranges were set based on recommendations from Dr. Steven Chapra (Stantec Consulting 2010) and from Bowie and colleagues (1985) (Table 16.9). Within the autocalibration, a fitness statistic is evaluated for each state variable as the reciprocal of a weighted average of the normalized RMSE and estimated as follows:

$$f(x) = \left[\sum_{i=1}^{q} w_i\right] \left[\sum_{i=1}^{q} \frac{1}{w_i} \left[\frac{\frac{1}{m}\sum_{j=1}^{m} O_{i,j}}{\left[\frac{1}{m}\sum_{j=1}^{m}\left(P_{i,j} - O_{i,j}\right)^2\right]^{1/2}} \right]\right] \quad \textbf{(21)}$$

Where,

$O_{i,j}$ = observed value

$P_{i,j}$ = predicted value

m = number of pairs of predicted and observed values

w_i = weighting factor

q = number of different state variables (e.g., dissolved oxygen, pH) in a bounded *n*-dimensional space for $x \equiv (x_1, x_2, ..., x_n) \quad x_k \in [0.0, 1.0]$ (Pelletier et al. 2006).

Table 16.9. Appropriate ranges (Min Value and Max Value) of parameters for QUAL2Kw modeling with the "Value" column showing the default value used. The "Auto-Cal" column indicates whether a parameter was autocalibrated in the 2010 modeling efforts.

Parameter	Value	Units	Symbol	Autocalibration inputs		
				Auto-cal	Min Value	Max Value
Stoichiometry						
Carbon	40	gC	gC	No	30	60
Nitrogen	7.2	gN	gN	No	5	9
Phosphorus	1	gP	gP	No	0.5	2
Dry weight	100	gD	gD	No	100	100
Chlorophyll	1	gA	gA	No	0.5	2
Inorganic suspended solids						
Settling velocity	Manual	m/d	v_i	No	0.2	2
Oxygen:						
Reaeration model	Manual Determination of Appropriate Formula			No		
Temp correction	1.024		q_a			
Reaeration wind effect	None					
O_2 for carbon oxidation	2.69	gO_2/gC	r_{oc}			

Parameter	Value	Units	Symbol	Autocalibration inputs		
				Auto-cal	Min Value	Max Value
O_2 for NH_4 nitrification	4.57	gO_2/gN	r_{on}			
Oxygen inhib model CBOD oxidation	Exponential					
Oxygen inhib parameter CBOD oxidation	0.60	L/mgO_2	K_{socf}	No	0.60	0.60
Oxygen inhib model nitrification	Exponential					
Oxygen inhib parameter nitrification	0.60	L/mgO_2	K_{sona}	No	0.60	0.60
Oxygen enhance model denitrification	Exponential					
Oxygen enhance parameter denitrification	0.60	L/mgO_2	K_{sodn}	No	0.60	0.60
Oxygen inhib model phyto resp	Exponential					
Oxygen inhib parameter phyto resp	0.60	L/mgO_2	K_{sop}	No	0.60	0.60
Oxygen enhance model bot alg resp	Exponential					
Oxygen enhance parameter bot alg resp	0.60	L/mgO_2	K_{sob}	No	0.60	0.60
Slow Carbonaceous biochemical oxygen demand						
Hydrolysis rate	0	/d	k_{hc}	No	0.05	0.25
Temp correction	1.047		q_{hc}	No	1	1.07
Oxidation rate	0.103	/d	k_{dcs}	No	0.05	0.25
Temp correction	1.047		q_{dcs}	No	1	1.07
Fast Carbonaceous biochemical oxygen demand						
Oxidation rate	10	/d	k_{dc}	No	0	10
Temp correction	1.047		q_{dc}	No	1	1.07
Organic Nitrogen						
Hydrolysis		/d	k_{hn}	Yes	0.05	0.3
Temp correction	1.07		q_{hn}	No	1	1.07
Settling velocity		m/d	v_{on}	Yes	0.05	0.25
Ammonium						
Nitrification		/d	k_{na}	Yes	0.05	4
Temp correction	1.07		q_{na}	No	1	1.07
Nitrate:						
Denitrification		/d	k_{dn}	Yes	0.05	2
Temp correction	1.07		q_{dn}	No	1	1.07
Sed denitrification transfer coefficient		m/d	v_{di}	Yes	0	1
Temp correction	1.07		q_{di}	No	1	1.07
Organic Phosphorus						
Hydrolysis		/d	k_{hp}	Yes	0.05	0.3
Temp correction	1.07		q_{hp}	No	1	1.07
Settling velocity		m/d	v_{op}	Yes	0.05	0.25
Inorganic Phosphorus						

Parameter	Value	Units	Symbol	Autocalibration inputs		
				Auto-cal	Min Value	Max Value
Settling velocity		m/d	v_{ip}	Yes	0	2
Sed P oxygen attenuation half sat constant		mgO_2/L	k_{spi}	Yes	0	2
Phytoplankton						
Max Growth rate		/d	k_{gp}	Yes	1.5	3
Temp correction	1.07		q_{gp}	No	1	1.07
Respiration rate		/d	k_{rp}	Yes	0.05	0.5
Temp correction	1.07		q_{rp}	No	1	1.07
Death rate		/d	k_{dp}	Yes	0	1
Temp correction	1		q_{dp}	No	1	1.07
Nitrogen half sat constant	15	ugN/L	k_{sPp}	No	10	25
Phosphorus half sat constant	2	ugP/L	k_{sNp}	No	1	5
Inorganic carbon half sat constant	1.30E-05	moles/L	k_{sCp}	No	1.30E-06	1.30E-04
Phytoplankton use HCO_3- as substrate	Yes					
Light model	Smith					
Light constant	57.6	langleys /d	K_{Lp}	No	40	110
Ammonia preference	15	ugN/L	k_{hnxp}	No	15	30
Settling velocity		m/d	v_a	Yes	0.05	0.5
Bottom Plants						
Growth model	Zero-order					
Max Growth rate		$gD/m^2/d$ or /d	C_{gb}	Yes	1.5	200
Temp correction	1.07		q_{gb}	No	1	1.07
First-order model carrying capacity	100	gD/m^2	$a_{b,max}$	No	50	200
Basal respiration rate		/d	k_{r1b}	Yes	0.02	0.2
Photo-respiration rate parameter	0.39	unitless	k_{r2b}	No	0	0.6
Temp correction	1.07		q_{rb}	No	1	1.07
Excretion rate		/d	k_{eb}	Yes	0	0.5
Temp correction	1.07		q_{db}	No	1	1.07
Death rate		/d	k_{db}	Yes	0	5
Temp correction	1.07		q_{db}	No	1	1.07
External nitrogen half sat constant		ugN/L	k_{sPb}	Yes	100	500
External phosphorus half sat constant		ugP/L	k_{sNb}	Yes	25	100
Inorganic carbon half sat constant		moles/L	k_{sCb}	Yes	1.30E-06	1.30E-04
Bottom algae use HCO_3- as substrate	Yes					
Light model	Half saturation					

Parameter	Value	Units	Symbol	Autocalibration inputs		
				Auto-cal	Min Value	Max Value
Light constant		langleys/d	K_{Lb}	Yes	40	100
Ammonia preference		ugN/L	k_{hnxb}	Yes	15	30
Subsistence quota for nitrogen		mgN/gD	q_{0N}	Yes	0.36	1.44
Subsistence quota for phosphorus		mgP/gD	q_{0P}	Yes	0.05	0.2
Maximum uptake rate for nitrogen		mgN/gD/d	r_{mN}	Yes	350	1500
Maximum uptake rate for phosphorus		mgP/gD/d	r_{mP}	Yes	50	200
Internal nitrogen half sat ratio			$K_{qN,ratio}$	Yes	1.05	5
Internal phosphorus half sat ratio			$K_{qP,ratio}$	Yes	1.05	5
Nitrogen uptake water column fraction	1		$N_{UpWCfrac}$	No	0	1
Phosphorus uptake water column fraction	1		$P_{UpWCfrac}$	No	0	1
Detritus (Particulate Organic Matter)						
Dissolution rate		/d	k_{dt}	Yes	0.05	5
Temp correction	1.07		q_{dt}	No	1.07	1.07
Settling velocity	0.4033805	m/d	v_{dt}	Yes	0.05	0.5

This tool allows the coefficient of variation of the RMSE (model results versus observed data) between each constituent along with appropriate, individual weighting factors (Table 16.10), to be summarized in a single value that the genetic algorithm seeks to maximize by adjusting all desired parameters.

Table 16.10. Weighting factors for each constituent used to calculate the fitness in model calibration.

Parameter	Weighting Factor
DO (mgO_2/L)	5
CBODs (mgO_2/L)	1
Norg (ugN/L)	2
NH_4 (ugN/L)	3
NO_3 (ugN/L)	3
Porg (ugN/L)	2
Inorg P (ugP/L)	4
Phyto (ugA/L)	1
Alk ($mgCaCO_3$/L)	4
pH	4
TN (ugN/L)	3
TP (ugP/L)	3

Parameter	Weighting Factor
TSS (mgD/L)	1
CBODu (mgO$_2$/L)	1
DO (mgO$_2$/L) - Min	5
DO (mgO$_2$/L) - Max	5
CH-A - Min	1
CH-A - Max	1

The constituents included in the fitness statistic for the 2010 modeling efforts heavily weighted DO average, minimum, and maximum values at Station E as indicted by a weighting factor of 5 and were established via discussions with Greg Pelletier and Nick von Stackelberg. The preliminary calibration parameters for each study site were established by the autocalibration algorithm and are outlined within each model and the associated documentation delivered to UDEQ.

Model Validation and Corroboration

At two locations (Silver Creek and Fairview) validation or corroboration datasets (identical to the calibration datasets) were collected during a different time period. These datasets were collected to determine whether the model calibrations held during a different time period under somewhat different conditions. For model corroboration, the boundary condition, point inflow, and weather data were updated to coincide with the conditions during the validation time period. All other site-specific information (e.g., channel characteristics) and parameters set during calibration were held constant. The exception was SOD, which can change during the year due to the transfer of oxygen-demanding material into the study reach. The SOD value for the validation period was again estimated based on *ER–GPP* at station E.

Findings, Recommendations, and Suggested Future Work

In general, the models resulting from this study were found to be capable of meeting the diverse intended uses. A number of the models have already been foundational in developing WLAs, and they are currently in the process of being used to assist in statewide nutrient criteria development. However, given the generic nature of the data collection and automatic calibration methods necessary to meet these varied needs and applications, there is at times significant uncertainty in important mechanisms and, therefore, in predictions. In some circumstances, additional data collection efforts and sensitivity and uncertainty analyses will be necessary to ensure the appropriate confidence in model predictions. Recommendations and suggested future efforts have also been identified.

Data Collection

Due to logistical constraints DWQ generally needs to collect the data necessary for model parameterization over 2-3 days. One of the most important lessons learned from this effort was that this requires collection of a large number of samples. These data collection efforts focused on collecting data during conditions that were presumed to be steady state, which led to the assumption that many of the data types gathered would not vary significantly throughout each day (including flow and water quality). The exceptions to this assumption were temperature, DO, pH, specific conductance, and chlorophyll-*a* (chl-*a*), which were measured at small time increments over 3–4 days in an effort to get a good understanding of their daily variability. While the streams themselves and the conditions at Stations A and B were relatively stable during these late summer time periods, the conditions downstream of many of the WRFs were not. Based on these studies, stable conditions do not exist for many of these plants, and the loads are highly variable throughout the day. This becomes critical to consider when sampling and modeling effluent-dominated systems (e.g., Silver Creek, Moroni). This variability caused significant problems during model population and calibration due to samples often not representing the average conditions and resulting in very different values between days when grab samples were collected. The Washington Department of Ecology generally samples twice a day for two days in a row in a stream. They also use a 24-hour composite for two days from the WRF effluent. They also average three benthic algae samples at randomly selected sites where periphyton are present. A similar approach may be warranted within the State of Utah; however, the representativeness of this sampling regime should be investigated.

Samples taken at different locations often did not coincide with the samples taken at a calibration location. To illustrate some of the potential spatial-temporal disconnects, assume a sample is taken at Station E (calibration location) at 10:00 am and this corresponded to the WRF release at 8 am (i.e., there is a 2-hour travel time between the WRF and Station E). The sample then taken at the WRF for the modeling occurs at 10:30 am. This and the measured flow value from the WRF are then used to calculate the load within the model that gets decayed and transported downstream to Station E (the calibration location). In this example, if the WRF effluent varies significantly over short time periods it is clear how the samples used in model forcing and calibration can easily be disconnected and how this disconnect influences model calibration and interpretation. These sorts of issues could be addressed by taking more samples throughout the study period (e.g., three per day), which would provide a much better understanding of the mean and variance throughout the entire study period. However, the constituents requiring higher frequency sampling will likely be site specific and depend on the loads impacting the system (e.g., highly variable WRF versus a lagoon system). Due to the WRF variability, the two data points gathered often provided a large range of possible concentrations and made it difficult to decide on an appropriate representative average to be used in the load estimates or in calibration. Another concern was that many times the data points were missing, resulted in nondetects, etc. These missing data provided even less information for model population and/or calibration. In the future when dealing with effluent-dominated streams with highly variable loads, it would be more appropriate to gather time-variable data and use the newly developed version of QUAL2Kw that allows for nonsteady flow with a continuous simulation option over a 365-day simulation period.

For understanding loads, it is important to collect enough flow information to ensure an appropriate water balance throughout the study reach. Since the ability to predict accurate concentrations hinges on correct volumes, in cases where the inflows are variable or discharge measurements show variability, more measurements are necessary. This includes ensuring accurate flow estimates at each of the study sites throughout the study period and may require the use of various flow measurement methods (e.g., slug injections rather than velocity area methods). Good flow data provide information regarding the appropriateness of the steady flow assumption and also provide a more solid estimate of average flow conditions if variability is present. This again highlights the need to understand the variability in WRF effluent. To assist in these efforts and all load allocation decision making (e.g., TMDLs or National Pollutant Discharge Elimination System permits), it is recommended that the state requires WRFs to track subhourly effluent rates and provide them to the state on a quarterly basis. It may be worthwhile to also have the WRFs install a water quality sonde and track the effluent DO, temperature, specific conductance, and pH since these data provide information regarding the plant effluent concentration variability and potential plant upsets.

The other key concern identified within these data had to do with analytical methods and the associated errors. For example, the sCBOD method detection limits are 3 or 5 mg/L, depending on the lab. With low sCBOD in WRF effluent and many of the streams not being highly influenced by high BOD loads, calculation of the actual BOD loads (from the WRF or at the upstream boundary condition) was based on concentrations assumed to be one-half the method detection limit. While other, less biased techniques exist to handle these censored values (i.e., trimmed mean, Winsorized mean, Cohen's maximum likelihood method), they are incapable of handling cases where more than 25% of the data are censored. These types of assumptions lead to significant errors in loads and the resulting sCBOD predictions. This may become a significant enough issue that new analytical methods need to be developed. Similarly, issues with analytical error were seen when it came to estimating some constituent concentrations based on differences (e.g., Organic N, Organic P, and detritus). The various sources of sampling and analytical error can produce significant errors in model loads and model calibration. This was particularly important given the limited number of samples and, again, illustrates the need for additional sampling throughout the study period. A measure of VSS should be included in the sampling protocol to better estimate detritus concentrations. Detritus could then become part of the fitness statistic and used in calibration.

Model Population and Calibration

Many of the issues associated with data collection have an obvious link to the success of model calibration, the ability to minimize model uncertainty, and the model's utility in decision making. Other concerns were identified that were more specifically related to model population or calibration.

A key concern was the very short spatial scale over which data were collected. In an effort to minimize the influence of tributaries, withdrawals, etc. and to meet the needs associated with quantifying open water metabolism, data were collected over short reaches based on Equation 7. This equation provides an estimated reach length where one-half of the oxygen has exchanged with the atmosphere via reaeration (Grace and Imberger 2006). While these distances were appropriate for the metabolism estimates, the associated short travel times resulted in data showing that many of the chemical reactions had a minimal influence within the study reach. In other words, over these travel time scales many reactions had minimal impact on instream concentrations, which resulted in relatively insensitive parameters. To address this concern, additional datasets were gathered in summer 2011 and the reach lengths were extended as much as possible while avoiding the tributary inflows, diversions, etc. that would require even more extensive data collection. For most future modeling applications (with the exception of WLA analyses) it would be best to make reach lengths as long as possible for the modeling study and deploy DO sensors within the reach at the optimal lengths based on Equation 7. If the approach provided by Grace and Imberger (2006) is still used, maximizing the multiplier is suggested (use 3 instead of 0.693, increasing the importance of instream processes from 50% to 95%, $-\ln(.05) \approx 3$) to ensure longer study reaches for the modeling and maintain the ability to still use the one- or two-station metabolism methods.

Another key issue identified in QUAL2Kw modeling is the need to decrease the number of parameters that are autocalibrated. As discussed previously, when possible, parameters should be measured or estimated for the study site of interest. Some key rates that can be estimated include the following:

1. BOD decomposition rates (k_d) could be estimated for each of the WRF types (lagoons, oxidation ditch, membrane) in Utah.
2. Nitrification rates could be estimated for each study site.
3. Photosynthetic active radiation attenuation within the water column could be estimated given the importance of bottom algae in many of these systems.

The other key parameter, at least in some systems, is SOD. A method of estimating SOD using DO measurements and metabolism methods was established; however, further investigation should be done into the assumptions about the minimal influence of other oxygen-demanding reactions that are reflected by the in-situ DO measurements and the amount of autotrophic respiration. The application of these methods to all systems needs to be investigated.

When it comes to autocalibration, there is an obvious need to decrease the number of parameters and potentially develop narrower ranges to confine autocalibration estimates. In these model applications, over 30 parameters are being optimized. This number is extremely high, but without more information regarding which parameters are unimportant, it is not clear which should be dropped from the autocalibration. The phytoplankton parameters seem to be insensitive. However, the bottom plant predictions are very important in many Utah streams, and the sensitive bottom algae parameters should be included in autocalibration—but which ones are sensitive is unknown. Approximately 15 of the parameters being optimized are associated with bottom plant growth, and information regarding the spatial and temporal concentrations to be used in calibration is minimal. Benthic algae carbon, N, and P ratios were established within some streams to provide an understanding of autotrophic nutrient limitation and provide insight into the heterotrophic resource quality. These data could be useful in bottom algae parameter estimation; however, they show significant spatial and temporal variability in stoichiometry along study reaches. This presents additional challenges in developing the appropriate sampling approaches to collect representative data at the reach or subreach scales. The utility of these data types and sampling techniques need to be further investigated. Additionally, the number of simulation days influence the bottom plant concentrations, and guidance regarding how best to set the simulation time period should be developed.

A sensitivity analysis should be completed for all these case studies to determine whether the number of autocalibrated parameters can be globally reduced (meaning for all study sites); if they can be reduced, then the parameters that are insensitive in different systems could be identified, reducing future data collection requirements. Within this effort, it would be important to identify which output parameters are important and influence the fitness statistic since the objective function (i.e., fitness) guides the calibration. If output values are not sensitive, they can influence the calibration algorithm performance.

In these applications, given the number of calibration parameters, short travel times, and limited amount of data, the resulting calibrations may or may not be appropriate for different circumstances. Some consistent findings suggest that key processes are missing, some parameters included in calibration are insensitive, or the approach to autocalibration may need some refinement. For example, the autocalibration algorithm consistently set sediment denitrification rates and inorganic P settling rates to relatively high values (Table 16.11). Both parameters provide a way to remove N and P from the water column, but in general this is done in a way that does not provide any insight into underlying mechanisms. In other words, these model terms are merely a sink for N and P. It is recommended that the influence of these parameters in autocalibration be investigated. If these additional N and P sinks truly exist, there is a need to investigate which mechanisms are not present within the model but are being consistently observed in these systems. Another interesting result of these calibrations are predicted pH values that are consistently too high. This can be important in the ability to predict other constituent concentrations, and the mechanisms leading to this should be revisited.

Table 16.11. Range of parameters found within the nine QUAL2Kw models within Utah.

Parameter	Brigham City	Fairview	Moroni	Oakley City	Price	Silver Creek	Spanish Fork	Tremon-ton	Wellsville	Avg.	Min.	Max.
Fitness	4.2	3.9	3.9	6.0	6.3	12.0	3.3	6.0	9.9			
Inorganic Suspended Solids												
Settling velocity	0.001	2	1.5	0.001	0.2	2	0.2	0.2	0.2	0.700	0.001	2.000
Oxygen												
Reaeration model	Internal	Tsivoglou-Neal	Owens-Gibbs	Internal	USGS (pool-riffle)	Tsivoglou-Neal	USGS (channel-control)	Internal	Owens-Gibbs			
Slow Carbonaceous Biochemical Oxygen Demand												
Oxidation rate	0.23	0.10	0.10	0.19	0.10	0.10	0.10	0.10	0.10	0.13	0.10	0.23
Organic Nitrogen												
Hydrolysis	0.08	0.30	0.26	0.22	0.25	0.08	0.25	0.27	0.09	0.20	0.08	0.30
Settling velocity	0.13	0.23	0.21	0.19	0.19	0.11	0.07	0.22	0.15	0.17	0.07	0.23
Ammonium												
Nitrification	3.60	4.00	1.74	3.48	0.05	3.10	3.84	0.93	0.87	2.40	0.05	4.00
Nitrate												
Denitrification	1.94	1.06	1.44	0.10	0.31	0.89	0.44	1.01	0.89	0.90	0.10	1.94
Sed denitrification transfer coeff	0.32	0.04	0.98	0.15	0.74	0.99	0.89	0.03	0.56	0.52	0.03	0.99
Organic Phosphorus												
Hydrolysis	0.10	0.09	0.08	0.15	0.13	0.12	0.11	0.28	0.24	0.15	0.08	0.28
Settling velocity	0.23	0.05	0.20	0.11	0.13	0.08	0.15	0.10	0.11	0.13	0.05	0.23
Inorganic Phosphorus												
Settling velocity	0.07	1.26	1.97	1.90	1.95	1.82	1.50	0.09	0.61	1.24	0.07	1.97
Sed P oxygen attenuation half sat constant	1.21	1.41	0.47	0.62	0.10	1.35	1.23	2.00	1.56	1.11	0.10	2.00
Phytoplankton												
Max Growth rate	2.2	2.0	2.8	1.8	2.4	2.9	2.8	1.8	2.9	2.4	1.8	2.9
Respiration rate	0.2	0.3	0.2	0.1	0.2	0.4	0.2	0.1	0.1	0.2	0.1	0.4
Death rate	0.4	0.6	0.2	0.7	0.1	0.7	0.8	0.0	0.1	0.4	0.0	0.8
Ammonia preference	15.0	25.3	26.2	15.0	16.8	19.7	16.2	23.2	19.9	19.7	15.0	26.2
Settling velocity	0.1	0.4	0.4	0.4	0.1	0.3	0.2	0.1	0.1	0.2	0.1	0.4
Bottom Plants												
Max Growth rate	10.2	39.5	15.7	85.0	15.8	60.6	39.2	161.1	8.6	48.4	8.6	161.1
Basal respiration rate	0.15	0.19	0.11	0.07	0.07	0.20	0.20	0.05	0.09	0.13	0.054	0.20
Photo-respiration rate parameter	0.39	0.01	0.01	0.39	0.01	0.01	0.01	0.01	0.01	0.09	0.010	0.39
Excretion rate	0.25	0.11	0.28	0.12	0.33	0.07	0.00	0.36	0.39	0.21	0.003	0.39
Death rate	0.65	0.07	2.64	0.01	1.67	0.01	0.01	4.46	4.03	1.50	0.005	4.46
External N half sat constant	389	253	374	264	350	180	465	320	184	309	180	465
External P half sat constant	47	68	48	63	67	76	56	57	90	64	47	90

Parameter	Brigham City	Fairview	Moroni	Oakley City	Price	Silver Creek	Spanish Fork	Tremon-ton	Wellsville	Avg.	Min.	Max.
Inorganic carbon half sat constant	1.7E-05	9.1E-05	1.2E-04	1.1E-04	7.4E-05	3.4E-05	7.8E-05	9.0E-05	2.5E-05	0.000	0.000	0.000
Light model	Half saturation	Smith	Smith	Half saturation	Smith	Smith	Smith	Smith	Smith			
Light constant	55	66	64	87	69	57	48	46	55	61	46	87
Ammonia preference	16	26	26	15	18	23	23	30	17	21	15	30
Subsistence quota for N	1.4	0.7	0.9	1.0	0.9	1.0	0.8	0.7	1.4	1.0	0.7	1.4
Subsistence quota for P	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.2
Maximum uptake rate for N	481	431	427	1405	744	764	957	724	1056	776	427	1405
Maximum uptake rate for P	117	101	175	184	145	163	98	124	146	139	98	184
Internal N half sat ratio	1.8	1.2	1.6	4.4	1.6	4.4	3.5	1.5	2.1	2.5	1.2	4.4
Internal P half sat ratio	2.1	3.5	1.3	4.8	5.0	3.2	3.9	1.4	2.8	3.1	1.3	5.0
Detritus (Particulate Organic Matter)												
Dissolution rate	3.70	1.58	4.75	1.63	0.28	4.68	1.07	0.07	1.66	2.16	0.07	4.75
Temp correction	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07			
Settling velocity	0.07	0.42	0.22	0.16	0.07	0.16	0.49	0.11	0.48	0.24	0.07	0.49
User-defined Autocalibration Parameters (Optional)												
Prescribed SOD (g O_2/m^2/day)	0.0	0.0	17	0	0.1	11	0	0	5	3.7	0	17

Use of Models in Support of Nutrient Criteria Development

As mentioned previously, this work was part of a greater effort to provide information to guide the development of nutrient criteria for the state of Utah. The goal was to evaluate changes in ecosystem structure (fish and macroinvertebrate communities), ecosystem function (whole stream metabolism, nutrient limitation, organic matter storage, and decompositions rates), and water chemistry and quality above and below each of the treatment plant discharges. The proposed NNC will consist of N and P limits, as well as other response indicators of primary production, ecosystem composition, and ecosystem function (Figure 16.4).

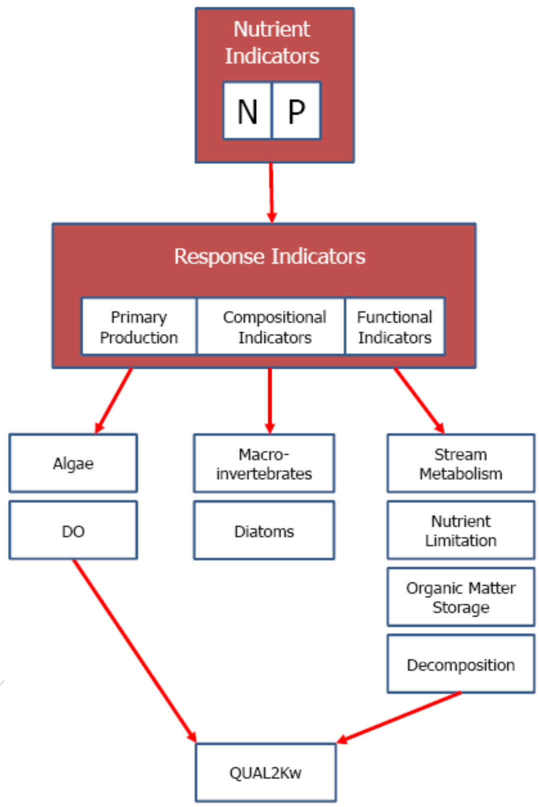


Figure 16.4. Numeric indicators of excess nitrogen and phosphorus pollution

Using the QUAL2Kw models built and calibrated for each study site, it is possible to predict the effects of nutrient addition, or removal, on these response variables. The model provides an additional line of evidence for the development of NNC by linking excess N and P levels in streams to thresholds in response variables such as algal growth and DO.

The QUAL2Kw models will be applied to nutrient criteria development using critical conditions for flow, meteorology, and water quality—either the same as the calibration conditions or similar to those generated for wasteload analyses (UDEQ, 2012). Within the model, inorganic N and P concentrations will be adjusted to identify the concentration that will result in just meeting the threshold level for each response indicator. Each nutrient will be analyzed separately (i.e., when conducting the P criteria analysis, N concentration will be set high enough so as not to limit algal growth).

The following are potential linkages between QUAL2Kw output and each response indicator (Figure 11.4):

1. Primary Productivity
   a. Benthic Algae (as measured by either chl-*a* or ash free dry mass): QUAL2Kw direct output, expressed as either chl-*a* or total algal biomass, will be compared to the recreation-based threshold of 150 mg/m^2 of chl-*a*.
   b. DO: QUAL2Kw direct output will be compared to existing water quality standards for DO with and without early life stages present.
   c. pH: QUAL2Kw direct output will be compared to maximum pH and diel change in pH.
2. Compositional Indicators: QUAL2Kw does not currently address compositional indicators.
3. Functional Indicators
   a. Stream Metabolism: QUAL2Kw direct output of GPP, expressed as g O_2/m^2, will be compared to thresholds developed by the ecological study.
   b. Nutrient Limitation: Determine whether N or P is the limiting nutrient at critical condition.
   c. Organic Matter Storage: Total organic matter storage in the sediments is not currently a standard output of QUAL2Kw. Typically, much of the organic matter gets deposited in the sediments outside of the simulation period and is added as prescribed SOD in the model.
   d. Decomposition Rate: Decomposition rate of organic matter does not currently vary by nutrient concentration in QUAL2Kw. Once the scientific literature quantifies the relationship between decomposition rate (both in the water column and sediments), researchers will inquire with Washington Department of Ecology whether this functionality can be incorporated into the model.

The study output is anticipated to be N and P criteria that meet the proposed response indicator thresholds for algal biomass, DO, pH, and GPP. By applying this approach to multiple models in different physiographic settings, a range of N and P criteria can be developed for stream and river systems in Utah.

Chapter 17

LITERATURE CITED

Acuna, V., A. Giorgi, I. Munoz, U. Uehlinger, and S. Sabater. 2004. Flow extremes and benthic organic matter shape the metabolism of a headwater Mediterranean stream. Freshwater Biology 49(7):960–971.

Allan, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35: 257–284.

Allgeier, J. E., A. D. Rosemond, and C. A. Layman. 2011. The frequency and magnitude of non-additive responses to multiple nutrient enrichment. Journal of Applied Ecology 48:821–834.

Appling, A. P., R. O. Hall, M. Arroita, and C. B. Yackulic. 2018. streamMetabolizer: Models for Estimating Aquatic Photosynthesis and Respiration. R package version 0.10.9. https://github.com/USGS-R/streamMetabolizer

Arbia, G., and G. Lafratta. 1997. Evaluating and updating the sample design in repeated environmental surveys: monitoring air quality in Padua. Journal of Agricultural, Biological, and Environmental Statistics 2:451–466.

Arheimer B.,and R. Liden. 2000. Nitrogen and phosphorus concentrations from agricultural catchments—influence of spatial and temporal variables. Journal of Hydrology 31: 140-159.

Aristegi, L., O. Izagirre, and A. Elosegi. 2009. Comparison of several methods to calculate reaeration in streams, and their effects of estimation of metabolism. Hydrobiologia 635:113–124.

Ascough, J. C., H. R. Maier, J. K. Ravalico, and M. W. Strudley. 2008. Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. Ecological Modeling 219:383–399.

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

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

Barbour, M. T., W. F. Swietlik, S. K. Jackson, D. L. Courtemanch, S. P. Davies, and C. O. Yoder. 2000. Measuring the attainment of biological integrity in the USA: A critical element of ecological integrity. Hydrobiologia 422:653–664.

Barko, J. W., and R. M. Smart, D. G. McFarkand, and R. L. Chen. 1981. Interrelationship between the growth of *Hydrilla verticillata* (L.f.) Royle and sediment nutrient availability. Aquatic Botany 32:205–216.

Barton, D. R., and S. M. Smith. 1984. Insects of extremely small and extremely large aquatic habitats. Pages 456–483 *in* V. H. Resh and D. M. Rosenberg (editors). The Ecology of Aquatic Insects. Praeger Publishing, New York.

Bennett, E. M., S. R. Carpenter, and N. F. Caraco. 2001. Human impact on erodable phosphorous and eutrophication: A global perspective, BioScience 51: 227–234.

Benstead, J. P., A. D. Rosemond, W. F. Cross, J. B. Wallace, S. L. Eggert, K. Suberkropp, V. Gulis, J. L. Greenwood, and C. J. Tant. 2009. Nutrient enrichment alters storage and fluxes of detritus in a headwater stream ecosystem. Ecology 90(9):2556–2566.

Bernhardt, E. S., and G. E. Likens. 2004. Controls on periphyton biomass in heterotrophic streams. Freshwater Biology 49:14–27.

Bernhardt E. S., G. E. Likens, R. O. Hall, D. C. Buso, S G. Fisher, T. M. Burton, J. L. Meyer, W. H. McDowell, M. S. Mayer, W. B. Bowden, and S. E. Findlay. 2005. Can't see the forest for the stream? In-stream processing and terrestrial nitrogen exports. AIBS Bulletin 55(3):219-30.

Bernot M. J., and W. K. Dodds. 2005. Nitrogen retention, removal, and saturation in lotic ecosystems. Ecosystems 8(4):442-53.

Bernot, M. J., D. J. Sobota, R. O. Hall, P. J. Mulholland, W. K. Haapala, J. R. Webster, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, N. B. Grimm, S. K. Hamilton, S. L. Johnson, W. H. McDowell, J. L. Meyer, B. Peterson, G. C. Poole, H. M. Valett, J. R. Webster, C. Arango, J. J. Beaulieu, A. J. Burgin, C. L. Crenshaw, A. M. Helton, L. Johnson, J. Merriam, B. R. Nitankederlehner, J. M. O'Brien, J. D. Potter, R. W. Scheibley, and K. Wilson. 2010. Inter-regional comparison of land-use effects on stream metabolism. Freshwater Biology 55: 1874–1890.

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

Biggs B. J. 1995. The contribution of flood disturbance, catchment geology and land use to the habitat template of periphyton in stream ecosystems. Freshwater Biology 33(3):419-38.

Biggs B.J. 2000. New Zealand periphyton guideline. Detecting, Monitoring and Managing Enrichment of Streams. NIWA, Christchurch.

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

Bilby, R. E., and G. E. Likens. 1980. Importance of organic debris dams in the structure and function of stream ecosystems. Ecology 61(5):1107–1113.

Bischoff, J., P. Massaro, and J. Strom. 2010. Jewitts Creek dissolved oxygen TMDL, edited by C. R. O. O. Water, Wenck Associates, Inc.

Bonin, H. L, R. P. Griffiths, and B. A. Caldwell. 2000. Nutrient and microbiological characteristics of fine benthic organic matter in mountain streams. Journal of the North American Benthological Society 19(2):235–249.

Borchardt M. A. 1996. Nutrients. Pages 184–218 *in* R. J. Stevenson, M. L. Bothwell, and R. L. Lowe (editors). Algal Ecology: Freshwater Benthic Ecosystems. Academic Press, San Diego, California.

Bott, T. L., D. S. Montgomery, J. D. Newbold, D. B. Arscott, C. L. Dow, A. K. Aufdenkampe, J. K. Jackson, and L. A. Kaplan. 2006. Ecosystem metabolism in streams of the Catskill mountains (Delaware and Hudson River watersheds) and Lower Hudson Valley. Journal of the North American Benthological Society 25(4):1018–1044.

Bouwman, A. F., G. Van Drecht, J. M. Knoop, A. H. W. Beusen, and C. R. Meinardi. 2005. Exploring changes in river nitrogen export the world's oceans. Global Biogeochemical Cycles 19(1):1–14.

Bowden, W. B., J. M. Glime, and T. Riis. 2006. Macrophytes and Bryophytes. Pages 381–414 *in* F. R. Hauer and G. A. Lamberti (editors). Methods in Stream Ecology. Elsevier/Academic Press, San Diego, California.

Bowie, G. L., W. B. Mills, D. B. Porcella, C. L. Campbell, J. R. Pagenkopf, G. L. Rupp, K. M. Johnson, P. W. H. Chan, S. A. Gherini, and C. E. Chamberlin. 1985. Rates, Constants, and Kinetic Formulations in Surface Water Quality Modeling. Edited by E. R. Laboratory, United States Environmental Protection Agency, Athens, Georgia.

Box, G. E. P., and N. R. Draper. 1987. Empirical Model-Building and Response Surfaces. Wiley Series in Probability and Statistics. Wiley, New York, New York.

Breiman L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and Regression Trees. The Wadsworth Statistics and Probability Series. Wadsworth International Group, Belmont, California.

Breiman, L. 1996. Bagging predictors. Machine Learning 24(2):123–140.

Breiman, L. 2001. Random Forests. Machine Learning 45(1):5–32.

Brooks, A. J., B. C. Chessman, and T. Haeusler. 2011. Macroinvertebrate traits distinguish unregulated rivers subject to water abstraction. Journal of the North American Benthological Society 30:419–435.

Brown, H. P. 1972. Trials and Tribulations of a Riffle Beetle Buff (Or Why Didn't I Stick with the Protozoa?). Bios 43 (2): 51-60.

Bunn S. E., P. M. Davies, and T. D. Mosisch. 1999. Ecosystem measures of river health and their response to riparian and catchment degradation. Freshwater Biology 41(2):333-45.

Bunn S. E., and P. M. Davies. 2000. Biological processes in running waters and their implications for the assessment of ecological integrity. Hydrobiologia 422(0): 61-70.

Burdon, F. J., A. R. McIntosh, and J. S. Harding. 2013. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. Ecological Applications 23:1036–1047.

Busse L. B., J. C. Simpson, and S. D. Cooper. 2006. Relationships among nutrients, algae, and land use in urbanized southern California streams. Canadian Journal of Fisheries and Aquatic Sciences. 63(12):2621-38.

Carlisle D. M., J. Falcone, M. R. Meador. 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental Monitoring and Assessment 151(1-4):143-60.

Capps, K. A., M. T. Booth, S. M. Collins, M. A. Davison, J. M. Moslemi, R. W. El-Sabaawi, J. L. Simonis, and A. S. Flecker. 2011. Nutrient diffusing substrata: a field comparison of commonly used methods to assess nutrient limitation. Journal of the North American Benthological Society 30(2):522–532.

Carr, J. F. 1962. Dissolved oxygen in Lake Erie, past and present. Publications of the Great Lakes Research Division, University of Michigan, 9:1–14.

Carr, G. M., A. Morin, A., and P. A. Chambers. 2005. Bacteria and algae in stream periphyton along a nutrient gradient. Freshwater Biology 50(8):1337–1350.

CH2M Hill Inc. 2012. Study of the Statewide Benefits of Nutrient Reduction. Utah Department of Environmental Quality, Division of Water Quality.

Chambers, P. A., E. E. Prepas, H. R. Hamilton, and M. L. Bothwell. 1991. Current Velocity and its effect on aquatic macrophytes in flowing waters. Ecological Applications 1(3):249–257.

Chao, C. T., and S. K. Thompson. 2001. Optimal adaptive selection of sampling sites. Environmetrics 12:517–538.

Chapra, S. C. 1997. Surface Water Quality Modeling, McGraw-Hill, Boston, Massachusetts.

Chapra, S.C., Pelletier, G.J. and Tao, H. 2008. QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality, Version 2.11. USA: Documentation and User's Manual. Civil and Environmental Engineering Department, Tufts University, Medford.

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(4):560-9.

Cho, J. H., and S. R. Ha. 2010. Parameter optimization of the QUAL2K model for a multiple-reach river using an influence coefficient algorithm. Science of The Total Environment 408(8):1985–1991.

Clarke, A., R. Mac Nally, N. Bond, and P. S. Lake. 2008. Macroinvertebrate diversity in headwater streams: a review. Freshwater Biology 53(9):1707–1721.

Clark, G. M., D. K. Mueller, and M. A. Mast. 2000. Nutrient concentrations and yields in undeveloped basins of the United States. Journal of the American Water Resources Association 36(4):849–860.

Connelly, N. M., M. R. Crossland, and R. G. Pearson. 2004. Effect of low dissolved oxygen on survival, emergence, and drift of tropical stream macroinvertebrates. Journal of the North American Benthological Society 23(2):251–270.

Correll D. L. 1998. The role of phosphorus in the eutrophication of receiving waters: a review. Journal of environmental quality 27(2):261-6.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Publications, Agencies and Staff of the U. S. Department of Commerce. Paper 189.

Cross, W. F., J. B. Wallace, A. D. Rosemond, and S. L. Eggert. 2006. Whole-system nutrient enrichment increase secondary production in a detritus-based ecosystem. Ecology 87:1556–1565.

Cuffney, T. F., M. E. Gurtz, and M. R. Meador. 1993. Methods for collecting benthic invertebrate samples as part of the Nation Water-Quality Assessment program: U.S. Geological Survey Open-File Report 93–406.

Cuffney, T. F., H. Zappia, E. M. P. Giddings, and J. F. Coles. 2005. Effects of urbanization on benthic macroinvertebraate assemblages in contrasting environmental settings: Boston, Massachusetts, Birmingham, Alabama and Salt Lake City, Utah. American Fisheries Symposium, 2005:361–407.

Cullen, A. C., and H. C. Frey. 1999. Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. Springer Science and Business Media.

Cummins K. W., R. C. Petersen, F. O. Howard, J. C. Wuycheck, and V. I. Holt. 1973. The utilization of leaf litter by stream detritivores. Ecology 54(2):336-345.

Dambacher, J. M., H. W. Li, and P. A. Rossignol. 2002. Relevance of community structure in assessing indeterminacy of ecological predictions. Ecology 83:1372–1385.

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251–66.

Dewson, Z. S., A. B. W. James, and R. G. Death. 2007. A review of the consequences of decreased flow for instream habitat and macroinvertebrates. Journal of the North American Benthological Society 26:401–415.

Dillman, D. A. 2011. Mail and Internet surveys: The tailored design method—2007 update with new internet, visual, and mixed-mode guide. John Wiley & Sons, New York.

Division of Water Quality. 2010. Strategic Monitoring Plan: 2010–2020. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah. Available at: https://documents.deq.utah.gov/water-quality/monitoring-reporting/DWQ-2017-010581.pdf.

Division of Water Quality. 2011. Field data collection for QUAL2Kw model build and calibration. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah.

Division of Water Quality. 2016. Utah's Final 2016 Integrated Report. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah. Available at: https://documents.deq.utah.gov/water-quality/monitoring-reporting/integrated-report/DWQ-2017-004941.pdf.

Dodds W. K. 2003. The role of periphyton in phosphorus retention in shallow freshwater aquatic systems. Journal of Phycology 39(5):840-9.

Dodds W. K. 2006. Eutrophication and trophic state in rivers and streams. Limnology and Oceanography 51(1, part 2):671-680.

Dodds, W. K. 2007. Trophic state, eutrophication and nutrient criteria in streams. Trends in Ecology and Evolution 22(12):669–676.

Dodds, W. K., W. W. Bouska, J. L. Eitzmann, T. J. Pilger, K. L. Pitts, A. J. Riley, J. T. Schloesser, and D. J. Thornbrugh. 2008. Eutrophication of US freshwaters: analysis of potential economic damages. Environmental Science and Technology 43(1):12-19.

Dodds, W. K., and J. J. Cole. 2007. Expanding the concept of trophic state in aquatic ecosystems: it's not just the autotrophs. Aquatic Sciences 69(4):427-439.

Dodds, W. K., and D. A. Gudder. 1992. The ecology of *Cladophora*. Journal of Phycology 28(4):415-27.

Dodds W. K., and R.M. Oakes. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography 2(10):333-341.

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., 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., and E. B. Welch. 2000. Establishing nutrient criteria in streams. Journal of the North American Benthological Society 19(1):186–196.

Downes, B. J. 2010. Back to the future: little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems. Freshwater Biology 55:60–79.

Dudley, T., S. Cooper, and N. Hemphill. 1986. Effects of Macroalgae on a Stream Invertebrate Community. Journal of the North American Benthological Society 5(2):93-106.

Earl S. R., H. M. Valett, and J. R.Webster. 2006. Nitrogen saturation in stream ecosystems. Ecology. 87(12):3140-51.

Edmondson, W. T., G. C. Anderson, and D. R. Petersen. 1956. Artificial eutrophication of Lake Washington. Limnology and Oceanography 1:4753.

Edinger J. E. 2001. Waterbody hydrodynamic and water quality modeling. An Introductory Work Book on Three Dimensional Water Body Hydrodynamics and Water Quality Modeling Including Modeling Software, Users Manual, Example Applications, and Derivations of the Model Relationships.

Elser J. J., M. E. Bracken, E. E. Cleland, D. S. Gruner, W. S. Harpole, H. Hillebrand, J. T. Ngai, E. W. Seabloom, J. B. Shurin, and J. E. Smith. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecology Letters 10(12):1135-42.

Elser J. J., W. F. Fagan, R. F. Denno, D. R. Dobberfuhl, A. Folarin, A. Huberty, S. Interlandi, S. S. Kilham, E. McCauley, K. L. Schulz, and E. H. Siemann. 2000. Nutritional constraints in terrestrial and freshwater food webs. Nature 408(6812):578.

Elser, J. J., E. Marzolf, and C. R. Goldman. 1990. The roles of phosphorus and nitrogen in limiting phytoplankton growth in freshwaters: a review of experimental techniques. Canadian Journal of Fisheries and Aquatic Sciences 47:1468–1477.

Ensign S. H., and M. W. Doyle. 2006. Nutrient spiraling in streams and river networks. Journal of Geophysical Research: Biogeosciences. 111(G4).

Environmental Protection Division. 1989. The Amplified Long-Term BOD Test, Protocol/Procedures and Test Specifications, edited by D. o. N. Resources. Atlanta, Georgia.

Evans, G., and L. Jones. 2001. Economic impact of the 2000 red tide on Galveston County, Texas: A Case Study. College Station, Texas: Department of Agricultural Economics, Texas A&M University.

Evans-White, M. A., B. E. Haggard, and J. T. Scott. 2014. A review of nutrient criteria development in the United States. Journal of Environmental Quality 42(4):1002–1014.

Fairchild, G. W., R. L. Lowe, and W. B. Richardson. 1985. Algal periphyton growth on nutrient diffusing Substrates: an in situ bioassay. Ecology 66(2)465–472.

Fausch, K. D., J. Lyons, J. R. Karr, and P. L. Angermeier. 1990. Fish Communities as indicators of environmental degradation. American Fisheries Society Symposium 8:123–144.

Federov, V. V., and C. Nachtsheim. 1995. Optimal designs for time-dependent responses. Pages 3–13 *in* C. P. Kitsos and W. G. Muller (editors). Proceedings of MODA4. Physica-Verlag, Heidelberg, Germany.

Ferreira, V., V. Gulis, and M. A. Graça. 2006. Whole-stream nitrate addition affects litter decomposition and associated fungi but not invertebrates. Oecologia 149(4): 718–729.

Ferson, S. 1996. What Monte Carlo methods cannot do? Human and Ecological Risk Assessment 2:990–1007.

Ferson, S. 2002. RAMAS Risk Calc: Risk Assessment with Uncertain Numbers. Lewis Press, Boca Raton, Florida.

Ferson, S., and L. R. Ginzburg. 1996. Different methods are needed to propagate ignorance and variability. Reliability Engineering and Systems Safety 54:133–144.

Ferson, S., W. Root, and R. Kuhn. 1999. RAMAS Risk Calc: Risk Assessment with Uncertain Numbers. Applied Biomathematics, Setauket, NY.

Finlay J. C. 2011. Stream size and human influences on ecosystem production in river networks. Ecosphere.2(8):1-21.

Findlay, S. E. G., R. L. Sinsabaugh, W. V. Sobczack, and M. Hoostal. 2003. Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter. Limnology and Oceanography 48(4):1608–1617.

Fisher, S. G., and G. E. Likens. 1973. Energy flow in Bear Brook, New Hampshire: an integrative approach to stream ecosystem metabolism. Ecological Monographs 43(4):421–439.

Florida Department of Environmental Protection. 2012. Technical support document: development of numeric nutrient criteria for Florida lakes, spring vents and streams.

Flynn, K., and M. N. Suplee. 2011. Using a computer water quality model to derive numeric nutrient criteria: Lower Yellowstone River. Helena, Montana: Montana Department of Environmental Quality.

Folt, C. L., C. Y. Chen, M. V. Moore, and J. Burnaford. 1999. Synergism and antagonism among multiple stressors. Limnology and Oceanography 44:864–877.

France R. 2011. Leaves as “crackers”, biofilm as “peanut butter”: Exploratory use of stable isotopes as evidence for microbial pathways in detrital food webs. Oceanological and Hydrobiological studies 40(4):110-5.

Francoeur, S. N. 2001. Meta-analysis of lotic nutrient amendment experiments: detecting and quantifying subtle responses. Journal of the North American Benthological Society 20(3):358–368.

Francoeur S. N., B. J. Biggs, R. A. Smith, and R. L. Lowe. 1999. Nutrient limitation of algal biomass accrual in streams: seasonal patterns and a comparison of methods. Journal of the North American Benthological Society 18(2):242-60.

Freeman M.C. 1986. The role of nitrogen and phosphorus in the development of *Cladophora glomerata* (L.) Kutzing in the Manwatu River, New Zealand. Hydrobiologia 131(1): 23-30.

Freeman, M. C., C. M. Pringle, and C. R. Jackson. 2007. Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales. Journal of the American Water Resources Association 43(1): 5–14.

France R. L. 1995. Carbon-13 enrichment in benthic compared to planktonic algae: foodweb implications. Marine ecology progress series 124:307-12.

Frey D. G. 1977. Biological integrity of water: an historical approach. Pages 127-140 In: The integrity of water. Proceedings of a symposium. US Environmental Protection Agency. Washington, DC.

Frey, H. C. 1993. Separating variability and uncertainty in exposure assessment: motivations and methods. Paper No. 93-79.01. Proceedings of the 86th Annual meeting, Air and Waste Management Association, Pittsburgh, Pennsylvia.

Fritz, K. M., M. M. Gangloff, and J. W. Feminella. 2004. Habitat modification by the stream macrophyte *Justicia americana* and its effects on biota. Oecologia 140(3): 388–397.

Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. C. Cleveland, P. A. Green, E. A. Holland, and D. M. Karl. 2004. Nitrogen cycles: Past, present, and future. Biogeochemistry 70:153–226.

Goodman, S. 2008. A dirty dozen: twelve p-value misconceptions. Seminars in Hematology 45:135–140.

Gorte R.W., C H. Vincent, L. A. Hanson, and M. R. Rosenblum. 2012. Federal land ownership: overview and data. Congressional Research Service.

Grace, M. G., and S. Imberger. 2006. Stream metabolism: performing and interpreting measurements.

Grattan, R. M., and K. Suberkropp. 2001. Effects of nutrient enrichment on yellow poplar leaf decomposition and microbial activity in streams. Journal of the North American Benthological Society 20:33–43.

Greenwood, J. L., and A. D. Rosemond. 2005. Periphyton response to long-term nutrient enrichment in a shaded headwater stream. Canadian Journal of Fisheries and Aquatic Sciences 62(9):2033–2045.

Gregory R., L. Failing, M. Harstone, G. Long, T. McDaniels, and D. Ohlson. 2012. Structured Decision Making: A Practical Guide to Environmental Management Choices. Wiley-Blackwell, Hoboken, New Jersey.

Gray L. J., and J. V. Ward. 1979. Food habits of stream benthos at sites of differing food availability. American Midland Naturalist 1: 157-67.

Grimm N. B., and S. G. Fisher. 1989. Stability of periphyton and macroinvertebrates to disturbance by flash floods in a desert stream. Journal of the North American Benthological 8(4):293-307.

Groves R. M. 2006. Nonresponse rates and nonresponse bias in household surveys. Public opinion quarterly 70(5):646-75.

Gulis, V., and K. Suberkropp. 2002. Leaf litter decomposition and microbial activity in nutrient-enriched and unaltered reaches of a headwater stream. Freshwater Biology 48(1):123–134.

Haapala A., T. Muotka, and A. Markkola. 2001. Breakdown and macroinvertebrate and fungal colonization of alder, birch, and willow leaves in a boreal forest stream. Journal of the North American Benthological Society 20(3):395-407.

Hale, S. S., and J. F. Helthse. 2008. Signals from the benthos: development and evaluation of a benthic Index for the near shore gulf of Maine. Ecological Indicators 8:338–350.

Hall, R. O. 2011. Stream metabolism workshop. Utah State University, Logan Utah. Unpublished data.

Hall, R. O., and J. L. Meyer. 1998. The trophic significance of bacteria in a detritus-based stream food web. Ecology 79(6):1995–2012.

Hall, R. O., and J. L. Tank. 2003. Ecosystem metabolism controls nitrogen uptake in streams in Grand Teton National Park, Wyoming. Limnology and Oceanography 48(3):1120–1128.

Hall, R.O., J. B. Wallace, and S. L. Eggert. 2000. Organic matter flow in stream food webs with reduced detrital resource base. Ecology 81:3445–3463.

Haggard B. E., E. H. Stanley, and D. E. Storm. 2005. Nutrient retention in a point-source-enriched stream. Journal of the North American Benthological Society 24(1):29-47.

Hansen, R. A., D. D. Hart, and R. A. Merz. 1991. Flow mediates predator-prey interactions between triclad flatworms and larval blackflies (Diptera: Simuliidae). OIKOS 46:88–92.

Harpole, W. S., J. T. Hgai, E. W. Seabloom, E. T. Borer, M. E. Bracken, J. J. Elser, D. S. Gruner, H. Hillenbrand, J. B. Shurin, and J. E. Smith. 2011. Ecology Letters 14(9):963–62.

Hart, D. D., and C. M. Finelli. 1999. Physical-biological coupling in streams: the pervasive effects of flow on benthic organisms. Annual Review of Ecology and Systematics 30:363–395.

Hart, D. D., and C. T. Robinson. 1990. Resource limitation in a stream community: phosphorus enrichment effects on periphyton and grazers. Ecology 71:1494–1502.

Hauer, F. R., and G. Lamberti, editors. 2011. *Methods in stream ecology*. Elsevier/Academic Press, San Diego, CA.

Hawkins, C.P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10(5):1456–1477.

Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29(1):312–343.

Hayes, K. R., H. M. Regan, M. A. Burgman, R. H. Devlin, A. R. Kapusciniski, and J. M. Dambacher. 2006. Methods to address uncertainty in ecological risk assessment of genetically modified organisms.

Heatherly II, T. 2014. Acceptable nutrient concentrations in agriculturally dominant landscapes: A comparison of nutrient criteria approaches for Nebraska rivers and streams. Ecological indicators 45:355-63.

Heathwaite A. L. 2010. Multiple stressors on water availability at global to catchment scales: understanding human impact on nutrient cycles to protect water quality and water availability in the long term. Freshwater Biology 55: 241–257.

Hecky, R. E., and P. Kilham. 1988. Nutrient limitation of phytoplankton in freshwater and marine Environments: A review of recent evidence on the effects of enrichment. Limnology and Oceanography 33(4):796–822.

Hering D., R. K. Johnson, S. Kramm, S. Schmutz, K. Szoszkiewicz, and P. F. Verdonschot. 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. Freshwater Biology 51(9):1757-85.

Herlihy, A. T., S. G. Paulsen, J. V. Van Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. Journal of the North American Benthological Society 27(4):860–877.

Hill, W. R., and S. E. Fanta. 2008. Phosphorus and light colimit periphyton growth at subsaturating Irradiances. Freshwater Biology 53(2):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(1):368–380.

Hill, B. H., R. K. Hall, P. Husby, A. T. Herlihy, and M. Dunne. 2000. Interregional comparisons of sediment microbial respiration in streams. Freshwater Biology 44(3): 213–222.

Hoagland, P., D. M. Anderson, Y. Kaoru, and A. W. White. 2002. The economic effects of harmful algal blooms in the United States, assessment issues, and information needs. Estuaries 25(4b): 819–837.

Hobbs, R. J., J. E. Higgs, and J. A. Harris. 2006. Novel ecosystems: implications for conservation and restoration. Trends in Ecology and Evolution 24(11):559–605.

Holtgrieve, G. W., D. E. Schindler, T. A. Branch, and Z. A. A'mar. 2010. Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics. Limnology and Oceanography 55(3):1047–1063.

Horner, R. R., E. B. Welch, and R. B. Veenstra. 1983. Development of nuisance periphytic algae in laboratory streams in relation to enrichment and velocity. Pages 121–131 *in* R. G. Wetzel (editor). Periphyton of Freshwater Ecosystems. W. Junk Publishers, The Hague, The Netherlands.

Houser, J. N., P. J. Mulholland, and K. O. Maloney. 2005. Catchment disturbance and stream metabolism: patterns in ecosystem respiration and gross primary production along a gradient of upland soil and vegetation disturbance. Journal of the North American Benthological Society 24(3):538–552.

Howarth, R. W., and R. Marino. 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. Limnology and Oceanography 51(1, part 2):364–376.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

Izagirre, O., U. Agirre, M. Bermejo, J. Pozoz, and A. Elosegi. 2008. Environmental controls of whole stream metabolism identified from continuous monitoring of Basque streams. Journal of the North American Benthological Society 27(2):252–268.

Jakus, P.M., N. Nelson, and J. Ostermiller. 2017. Using survey data to determine a numeric criterion for nutrient pollution. *Water Resources Research*, *53*(12), pp.10188-10200.

Johnson R. K., D. Hering, M. T. Furse, and R. T. Clarke. 2006. Detection of ecological change using multiple organism groups: metrics and uncertainty. Hydrobiologia 566(1):115-37.

Johnson, L. T., J. L. Tank, and W. K. Dodds. 2009. The influence of land use on stream biofilm nutrient limitation across eight North American ecoregions. Canadian Journal of Fisheries and Aquatic Sciences 66:1081–1094.

Jones, J. B., J. D. Schade, S. G. Fisher, and N. B. Grimm. 1997. Organic matter dynamics in Sycamore Creek, a Desert Stream in Arizona, USA, Journal of North American Benthological Society 16(1): 78–82.

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(3):627–638.

Kail, J., J. Arel, and S. C. Jahnig. 2012. Limiting factors and thresholds for macroinvertebrate assemblages in European rivers: Empirical evidence from three datasets on water quality, catchment urbanization and river restoration. Ecological Indicators 18:63–72.

Kannel, P. R., S. Lee, Y. S. Lee, S. R. Kanel, and G. J. Pelletier. 2007. Application of automated QUAL2Kw for water quality modeling and management in the Bagmati River, Nepal., Ecological Modelling 202:503–517.

Kardouni, J., and N. Cristea. 2006. Snoqualmie River Temperature Total Maximum Daily Load Study, edited by Washington Department of Ecology, Olympia, Washington.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21–27.

Kelly, O., K. O. Maloney, and D. E. Weller. 2011. Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. Freshwater Biology 56: 611–626.

Kemp, M. J., and W. K. Dodds. 2001. Centimeter-scale patterns in dissolved oxygen and nitrification rate in a prairie stream. Journal of the North American Benthological Society 20(3):347–357.

Kiffney, P.M., J. S. Richardson, and J. P. Bull. 2004. Establishing light as a causal mechanism structuring stream communities in response to experimental manipulation of riparian buffer width. Journal of the North American Benthological Society *23*(3): 542-555.

King, R. S., and M. E. Baker. 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. Journal of the North American Benthological Society 29:998–1008.

King, R. S., M. E. Baker, P. F. Kazyak, and D. E. Weller. 2011. How novel is too novel? Stream community thresholds at exceptionally low levels of catchment urbanization. Ecological Applications 21:1659–1678.

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.

King, R. S., and C. J. Richardson. 2004. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. Environmental Management 31:795–809.

Kiry, P., D. Velinsky, and A. M. Compton. 1999. Determination of dry weight and percent organic matter for sediments, tissues, and benthic algae. PCER Procedure P-16-113. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia.

Kominoski, J. S., and A. D. Rosemond. 2012. Conservation from the bottom up: forecasting effects on global change on dynamics of organic matter and management needs for river networks. Freshwater Science 31(1):51–68.

Lancaster, J., and B. J. Downes. 2010. Linking the hydraulic world of individual organisms to ecological processes: putting ecology into ecohydraulics. River Research and Applications 26:385–403.

Lane, C. S., D. R. Lyon, and S. E. Ziegler. 2012. Cycling of two carbon substrates of contrasting lability by heterotrophic biofilms across a nutrient gradient of headwater streams. Aquatic Sciences 75(2): 235-250.

Le, N. D., and J. V. Zidek. 1994. Network designs for monitoring multivariate random spatial fields. Pages 191–206 *in* J. P. Vilaplana and M. L. Puri (editors). Recent Advances in Statistics and Probability. VSP, Leiden, The Netherlands.

Lee L. 2013. NADA: Nondetects and data analysis for environmental data. R package version15-6.

Levesque, V.A., and Oberg, K.A. 2012. Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3–A23, 148 p.

Lewis, S. L., and M. A. Maslin. 2015. Defining the anthropocene. Nature 519 (7542): 171.

Lewis, Jr, W. M., J. M. Melack, W. H. McDowell, M. McClain, and J. E. Richey. 1999. Nitrogen yields from undisturbed watersheds in the Americas. Biogeochemistry 46: 149–162.

Lewis, Jr., W. M., W. A. Wurtsbaugh, and H. W. Paerl. 2011. Rationale for control of anthropogenic Nitrogen and phosphorus to reduce eutrophication of inland waters. Environmental Science and Technology 45:10300–10305.

Liess, A., K. Lange, F. Schulz, J. J. Piggott, C. D. Matthaei, and C. R. Townsend. 2009. Light, nutrients and grazing interact to determine diatom species richness via changes to productivity, nutrient state and grazer activity. Journal of Ecology 97:326–336.

Link, W. A., E. Cam, J. D. Nichols, and E. G. Cooch. 2002. Of bugs and birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66:277–291.

Lohman K., J. R. Jones, and B. D. Perkins. 1992.Effects of nutrient enrichment and flood frequency on periphyton biomass in northern Ozark streams. Canadian Journal of Fisheries and Aquatic Sciences 149(6):1198-205.

Lovell, D. P. 2013. Biological importance and statistical significance. Journal of Agricultural and Food Chemistry 61:8340–8348.

Lowe, R. L., S. W. Golladay and J. R. Webster. 1986. Periphyton response to nutrient manipulation in streams draining clearcut and forested watersheds. Journal of the North American Benthological Society *5*(3): 221-229.

Lowe, W.H., and G. E. Likens. 2005. Moving headwater streams to the head of the class. AIBS Bulletin 55(3):196–197.

Macan, T. T. 1963. Freshwater Ecology. Longmans, Green and Co, London.

Madsen, T. V., and N. Cedergreen. 2002. Sources of nutrients to rooted submerged macrophytes growing in a nutrient-rich stream. Freshwater Biology 47:283–291.

Maine Department of Environmental Protection. 2009. Protocols for calculating the diatom total phosphorus index (DTPI) and diatom total nitrogen index (DTNI) for wadeable streams and rivers. DEPLW-0970A.

Maloney, K. O., and D. E. Weller. 2011. Anthropogenic disturbance and streams: land use and land use change affect stream ecosystems via multiple pathways. Freshwater Biology 56(3):611–626.

Marshall, B. D. 2001. An Evaluation of the Sensitivity of a Macroinvertebrate Biomonitoring Study in Headwater Streams of New River Gorge National River. Journal of Freshwater Ecology, 16:3, 415-428.

Marshall, B. D. 2008. Within site variation of composite Surber samples and single Surber Samples. In: Assessment of macroinvertebrate communities near natural gas development on the Pinedale Anticline Project Area (PAPA). Prepared for the Sublette County Conservation District, Pinedale WY.

Marshall, B. D. 2008. Assessment of macroinvertebrate communities near natural gas development on the Pinedale Anticline Project Area (PAPA). Prepared for the Sublette County Conservation District, Pinedale WY.

Marshall, B. D. 2009. Assessment of the biological condition of the New Fork River, in the vicinity of the Pinedale Anticline Project Area: 2008. Prepared for Sublette County Conservation District, Pinedale, Wyoming.

Martinez-Abrain, A. 2008. Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology. Acta Oecologica 34:9–11.

Mason, S. J., and G. E. Graham. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. Quarterly Journal of The Royal Meteorological Society 128:2145–2166.

Mazumder, A., and J. A. Edmonson. 2002. Impact of fertilization and stocking on trophic interactions and growth of juvenile sockeye salmon (*Oncorhynchus nerku*). Canadian Journal of Fisheries and Aquatic Sciences 59:1361–1373.

McBride, G. B., and S. Chapra. 2005. Rapid calculation of oxygen in streams: approximate delta method, Journal of Environmental Engineering a3a (3): 336–342.

McCormick P. V. , M. B. O'Dell, R. B. Shuford III, J. G. Backus, Kand W. C. Kennedy. Periphyton responses to experimental phosphorus enrichment in a subtropical wetland. Aquatic Botany 71(2):119-39.

McCune, B. 2011. Nonparametric multiplicative regression for habitat modeling.Oregon State University, OR.

McLaughlin, D. B. 2012. Assessing the predictive performance of risk-based water quality criteria using decision error estimates from receiver operating characteristics (ROC) analysis. Integrated Environmental Assessment and Management. 8(4):674–684.

McTammany, M. E., E. F. Benfield, and J. R. Webster. 2007. Recovery of stream ecosystem metabolism from historical agriculture. Journal of the North American Benthological Society 26(3):532–545.

Menzie C., M. H. Henning, J. Cura, K. Finkelstein, J. Gentile, J. Maughan, D. Mitchell, S. Petron, B. Potocki, S. Svirsky, P. Tyler. 1996. Special report of the Massachusetts weight-of-evidence workgroup A weight-of-evidence approach for evaluating ecological risks. Human and Ecological Risk Assessment 2(2):277-304.

Merow, C., M. J. Smith, T. C. Edwards, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wuest, N. E. Zimmermann, and J. Elith. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography doi: 10.1111/ecog.00845.

Meybeck, M. 2003. Global analysis of river systems: from Earth system controls to Anthropocene syndromes. Philosophical Transactions of the Royal Society of London B: Biological Sciences 358(1440):1935–1955.

Meyer, J. L., D. L. Strayer, J. B. Wallace, S. L. Eggert, G. S. Helfman, and N. E. Leonard. 2007. The contribution of headwater streams to biodiversity in river networks. Journal of the American Water Resources Association 43(1): 86–103.

Meyer, S. M. 2004. End of the wild. The extinction crisis is over. We lost. Boston Review, April/May.

Meyer J. L., D. L. Strayer, J. B. Wallace, S. L. Eggert, G. S. Helfman, and N. E. Leonard. 2007. The contribution of headwater streams to biodiversity in river networks1. JAWRA Journal of the American Water Resources Association 43(1):86-103.

Miltner, R. J., and A. E. T. Rankin. 1998. Primary nutrients and the biotic integrity of rivers and streams. Freshwater Biology 40(1):145–158.

Minshall, G. W. 1978. Autotrophy in streams. Bioscience 28(12):767–775.

Minshall, G. W. 1984. Aquatic insect-substratum relationships. Pages 358–400 *in* V. H. Resh and D. M. Rosenberg (editors). The Ecology of Aquatic Insects. Praeger Publishing, New York.

Minnesota Pollution Control Agency. 2005. Minnesota lake water quality assessment report: developing nutrient criteria, third edition.

Minnesota Pollution Control Agency. 2008. Relation of nutrient concentrations and biological responses in Minnesota streams: applications for river nutrient criteria development.

Mitsui, A., S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. 1986. Strategy by which nitrogen-fixing unicellular Cyanobacteria grow photoautotrophically. Nature 323: 720–722.

Montana Department of Environmental Quality. 2008. Scientific and technical basis of the numeric nutrient criteria for Montana.

Morgan, M. G., and M. Henrion. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, 6th edition. Cambridge University Press, Cambridge, United Kingdom.

Mulholland, P. J., C. S. Fellows, J. L. Tank, N. B. Grimm, J. R. Webster, S. K. Hamilton, E. Marti, 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 diurnal dissolved oxygen profiles as indicators of in-stream metabolism and disturbance effects: Fort Benning as a case study. Ecological Indicators 5:243–252.

Mulholland P. J., A. D. Steinman, E. R. Marzolf, D. R. Hart, and D. L. DeAngelis. 1994. Effect of periphyton biomass on hydraulic characteristics and nutrient cycling in streams. Oecologia 98(1):40-7.

Mulholland, P. J., A. D. Stienman, A. V. Palumbo, and D. L. DeAngelis. 1991. Journal of the North American Benthological Society 10(2):127–142.

Muotka, T., and P. Laasonen. 2002. Ecosystem recovery in restored headwater streams: the role of enhanced leaf retention. Journal of Applied Ecology 39(1):145–156.

Naeem, S. 2002. Ecosystem consequences of biodiversity loss: the evolution of a paradigm. Ecology83(6): 1537-1552.

Neilson, B. T., A. J. Hobson, N. von Stackelberg, M. P. Shupryt, and J. D. Ostermiller. 2012. Using Qual2k modeling to support nutrient criteria development and wasteload analysis in Utah. Technical Report to the Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah.

Nelson N. M., J. B. Loomis, P. M. Jakus, M. J. Kealy, N. von Stackelburg, and J. Ostermiller. 2015. Linking ecological data and economics to estimate the total economic value of improving water quality by reducing nutrients. Ecological Economics 118:1-9.

Nevers, M. B., and R. L. Whitman. 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. Water Research 45(4):1659–1668.

Newbold J. D., J. W. Elwood, R. V. O'Neill, and W. V. Winkle. 1981. Measuring nutrient spiraling in streams. Canadian Journal of Fisheries and Aquatic Sciences 38(7):860-3.

Newbury, R. W. 1984. Hydrological determinants of aquatic insect habitats. Pages 323–357 *in* V. H. Resh and D. M. Rosenberg (editors). The Ecology of Aquatic Insects. Praeger Publishing, New York.

Norris R. H., and C. P. Hawkins. 2000. Monitoring river health. Hydrobiologia 435(1-3):5-17.

Norton D. J., J. D. Wickham, T. G. Wade, K. Kunert, J. V. Thomas, P. Zeph. 2009. A method for comparative analysis of recovery potential in impaired waters restoration planning. Environmental Management 44(2):356-68.

Nydick, K. R., B. Moraska, J. Lafrancois, S. Baron, and B.M. Johnson. 2004. Nitrogen regulation of algal biomass, productivity, and composition in shallow mountain lakes, Snowy Range, Wyoming, USA. Canadian. Journal Fisheries and Aquatic Sciences 61: 1256–1268.

O'Brien J. M, W. K. Dodds. 2010. Saturation of NO_3^- uptake in prairie streams as a function of acute and chronic N exposure. Journal of the North American Benthological Society 29(2):627-35.

Odum, H. T. 1956. Primary production in flowing waters. Limnology and Oceanography 1:103–117.

Olsen. A. R., and D. V. Peck. 2008. Survey design and extent estimates for the wadeable streams assessment. Journal of the North American Benthological Society 27(4):822–836.

Olson, J. R., and C. P. Hawkins. 2013. Developing site-specific nutrient criteria from empirical models. Freshwater Science 32(3):719–740.

Omernik, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American geographers 77(1):118-125.

Opsahl, R. W., T. W. Wellnitz, and N. L. Poff. 2003. Current velocity and invertebrate grazing regulate Stream algae: results of an in situ electrical exclusion. Hydrobiologia 499:135–145.

Palmer S. M., D. Hope, M. F. Billett, J. J. Dawson, and C. L. Bryant. 2001. Sources of organic and inorganic carbon in a headwater stream: evidence from carbon isotope studies. Biogeochemistry 52(3):321-38.

Paasy, S. I. 2008. Continental diatom biodiversity in stream benthos declines as more nutrients become limiting. Proceedings of the National Academy of Sciences 105:9663–9667.

Paul, M. J. 2009. Development of models to establish links between human-caused nutrient enrichment and alterations to the composition of stream communities. Report to USEPA, Region 8.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. Annual review of Ecology and Systematics 32(1):333–365.

Paulsen, S. G., A. Mayio, D. V. Peck, J. L. Stoddard, E. Tarquinio, S. M. Holdsworth, J. Van Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufman, M. T. Barbour, D. P. Larsen, and A. R. Olsen. 2008. Condition of stream ecosystems in the US: and overview of the first national assessment. Journal of the North American Benthological Society 27(4):812–821.

Pelletier, G. J., and S. Chapra. 2008. QUAL2Kw theory and documentation (version 5.1): A modeling framework for simulating river and stream water quality. Edited by Washington State Department of Ecology, Olympia, Washington.

Pelletier, G. J., S. C. Chapra, and H. Tao. 2006. QUAL2Kw—A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration., Environmental Modelling & Software 21(3), 419–425.

Peterson, B. J., L. Deegan, J. Helfrich, J. E. Hobbie, M. Hullar, B. Moller, T. E. Ford, A. Hershey, A. Hiltner, G. Kipphut, M. A. Lock, D. M. Fiebig, V. McKinley, M. C. Miller, J. R. Vestal, R. Ventullo, and G. Volk. 1993. Biological responses of a tundra river to fertilization. Ecology 74:653–672.

Pringle, C. M., P. D. Paaby-Hansen, P. D. Vaux, and C. R. Goldman. 1986. In situ assays of periphyto growth in a lowland Costa Rican stream. Hydrobiologia 134:207–213.

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

Quinn, G., and M. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press.

R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna, Austria. www/R-project.org.

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

Regan, H. M., Y. Ben-Haim, B. Langford, W. G. Wilson, P. Lundberg, and S. J. Andelman. 2005. Robust decision making under severe uncertainty for conservation management. Ecological Applications 15(4):1471–1477.

Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications 12(2):618–628.

Richards, D. C., T. Arrington, and W. VanWinkle. 2009. Metapopulation viability analysis and quantitative risk assessment of threatened snail in the Snake River, ID. Idaho Power Report to Federal Energy Regulatory Commission.

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

Riis T., W. K. Dodds, P. B. Kristensen, and A. J. Baisner. 2012. Nitrogen cycling and dynamics in a macrophyte-rich stream as determined by a release. Freshwater Biology 57(8):1579-91.

Riseng C. M., M. J. Wiley, and R. J. Stevenson. 2004. Hydrologic disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. Journal of the North American Benthological Society 23(2):309-26.

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, and M. Muller. 2011. pROC: An open source package for R and S+ to analyze and compare ROC curves. BMC Biometrics 12:77

Robinson, C. T., and M. P. Gessner. 2000. Nutrient addition accelerates leaf breakdown in an alpine spring brook. Oceologia 122:258–263.

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

Rosemond, A. D., P. J. Mulholland, and S. H. Brawley. 2000. Seasonally shifting limitation of stream periphyton: Response of algal populations and assemblage biomass and productivity to variation in light, nutrients, and herbivores. Canadian Journal of Fisheries and Aquatic Sciences 57:66–75.

Sabater, S., S. V. Gregory, and J. R. Sedell. 2002. Community dynamics and metabolism of benthic algae colonizing wood and rock substrata in a forested stream. Journal of Phycology 34(4):561–567.

Sanderson, B. L., H. J. Coe, C. D. Tran, K. H. Macneale, D. L. Harstad, and A. B. Goodwin. 2009. Nutrient Limitation of periphyton in Idaho streams: results from nutrient diffusing substrate experiments. Journal of the North American Benthological Society 28(4):832–845.

Schiller, D.V., E. Martí, J. L. Riera and F. Sabater. 2007. Effects of nutrients and light on periphyton biomass and nitrogen uptake in Mediterranean streams with contrasting land uses. Freshwater Biology *52*(5): 891-906.

Schindler, D. W. 1977. Evolution of phosphorus limitation in lakes. Science 195:260–262.

Schindler, D. W., F. A. J. Armstrong, S. K. Holmgren, and G. J. Brunskill. 1971. Eutrophication of Lake 227, Experimental Lakes Area, Northwestern Ontario, by Addition of Phosphate and Nitrate. Journal of the Fisheries Research Board of Canada 28(11):1763–1782.

Scott, J. T., J. A. Back, J. M. Taylor, and R. S. King. 2008. Does nutrient enrichment decouple algal–bacterial production in periphyton? Journal of the North American Benthological Society 27(2):332–344.

Scott, J. H., D. A. Lang, R. S. King, and R. D. Doyle. 2009. Nitrogen fixation and phosphatase activity in periphyton growing on nutrient diffusing substrata: evidence for differential nutrient limitation in stream periphyton. Journal of the North American Benthological Society. 28(1):57–68.

Seife, C. 2010. Proofiness: the Dark Arts of Mathematical Deception. Penguin Books, New York, New York.

Seitzinger, S. P., E. Mayorga, A. F. Bouwman, C. Kroeze, A. H. Beusen, G. Billen, G. Van Drecht, E. Dumont, B. M. Fekete, J. Garnier, and J. A. Harrison. 2010. Global river nutrient export: A scenario analysis of past and future trends. Global Biogeochemical Cycles 24(4).

Slavik, K., B. J. Peterson, L. A. Deegan, W. B. Bowden, A. E. Hershey, and J. E. Hobbie. 2004. Long-term responses of the Kuparuk river ecosystem to phosphorus enrichment. Ecology 85:939–95.

Smith V. H. 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research 10(2):126-39.

Smith, R.A., R. B. Alexander, and G. E. Schwarz. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. Environmental Science and Technology 37: 3039-3047.

Smith, A. J., R. W. Bode, and G. S. Kleppel. 2007. A nutrient biotic index (NBI) for use with benthic macroinvertebrate communities. Ecological Indicators 7(2):371–386.

Smith A. J., B. T. Duffy, and M. A. Novak. 2015. Observer rating of recreational use in wadeable streams of New York State, USA: implications for nutrient criteria development. Water research 69:195-209.

Smith, V. H., and D. W. Schindler. 2009. Eutrophication science: where do we go from here? Trends in Ecology and Evolution 24:201–207.

Smith V. H., G. D. Tilman, and J. C. Nekola. 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environmental Pollution 100(1-3):179-96.

Smith A. J., and C. P. Tran. 2010. A weight-of-evidence approach to define nutrient criteria protective of aquatic life in large rivers. Journal of the North American Benthological Society 29(3):875-91.

Stantec Consulting. 2010. Jordan River TMDL: 2010 QUAL2Kw Model Calibration Technical Memo PUBLIC DRAFT. , edited by D. o. W. Q. Utah Department of Environmental Quality, Salt Lake City, UT.

Statzner, B., and L. A. Beche. 2010. Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? Freshwater Biology 55: 80–119.

Stelzer, R. S., J. Heffernan, and G. E. Likens. 2003. The influence of dissolved nutrients and particulate organic matter quality on microbial respiration and biomass in a forest stream. Freshwater Biology 48:1925–1937.

Stevenson R. J. 1990. Benthic algal community dynamics in a stream during and after a spate. Journal of the North American Benthological Society 9(3):277-88.

Stevenson R. J., B. J. Bennett, D. N. Jordan, and R. D. French. 2012. Phosphorus regulates stream injury by filamentous green algae, DO, and pH with thresholds in responses. Hydrobiologia 695(1):25-42.

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 561:140–165.

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

Stockner J. G., and K. R. Shortreed. 1978. Enhancement of autotrophic production by nutrient addition in a coastal rainforest stream on Vancouver Island. Journal of the Fisheries Board of Canada 35(1):28-34.

Stockner J. G., and E. A. Macisaac 1996. British Columbia lake enrichment programme: two decades of habitat enhancement for sockeye salmon. Regulated Rivers: Research & Management 12(4–5):547-61.

Stutter M. I., S. J. Langan, and R. J. Cooper. 2008. Spatial and temporal dynamics of stream water particulate and dissolved N, P and C forms along a catchment transect, NE Scotland. Journal of Hydrology 350(3-4):187-202.

Suberkropp K., V. Gulis, A. D. Rosemond, and J. P. Benstead. 2010. Ecosystem and physiological scales of microbial responses to nutrients in a detritus-based stream: Results of a 5-year continuous enrichment. Limnology and Oceanography 55(1):149-60.

Suplee M. W., A. Varghese, and J. Cleland. 2007. Developing Nutrient Criteria for Streams: An Evaluation of the Frequency Distribution Method. Journal of the American Water Resources Association 43(2):453-72.

Suplee M. W., and V. Watson. 2013. Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers—Update 1. Helena, MT: Montana Dept. of Environmental Quality.

Suplee M. W., V. Watson, W. K. Dodds, and C. Shirley. 2012. Response of algal biomass to large-scale nutrient controls in the Clark Fork River, Montana, United States. Journal of the American Water Resources Association 48(5):1008-21.

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 streams. Journal of the American Water Resources Association 45(1): 134–140.

Tank, J. L., M. J. Bernot, and E. J. Rosi-Marshall. 2006. Nitrogen limitation and uptake. Pages 213–238 *in* F. R. Hauer and G. A. Lamberti (editors). Methods in Stream Ecology. Elsevier, New York.

Tank, J. L., and W. K. Dodds. 2003. Nutrient Limitation of epilithic and epixylic biofilms in ten North American streams. Freshwater Biology 48:1031–1049.

Tank, J. L., E. J. Rosi-Marshall, N. A. Griffiths, S. A. Entrekin, and M. L. Stephen. 2010. A review of allochthonous organic matter dynamics and metabolism in streams. Journal of the North American Benthological Society 29(1):118–146.

Tank, J. L., and J. R. Webster. 1998. Interaction of substrate and nutrient availability on wood biofilm processes in streams. Ecology 79:2168–2179.

Thompson, S. K., and A. F. Seber. 1996. Adaptive Sampling. Wiley, New York, New York.

Townsend C. R. 1989. The patch dynamics concept of stream community ecology. Journal of the North American Benthological Society 8(1):36-50.

Townsend, C. R., and A.G. Hildrew. 1994. Species traits in relation to a habitat template for river systems. Freshwater Biology 31: 265–275.

Townsend, C. R., S. S. Uhlmann, and C. D. Matthaei. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810–1819.

Triska, F. J., and J. R. Sedell. 1976. Decomposition of four species of leaf litter in response to nitrate manipulation. Ecology 57:783–792.

U.S. Environmental Protection Agency. 1976. Quality criteria for water. EPA-440/9-76-023. Washington, DC.

U.S. Environmental Protection Agency. 1986. Quality Criteria for Water, Office of Water, EPA-440/5-86-001, Washington, DC.

U.S. Environmental Protection Agency. 2000. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

U.S. Environmental Protection Agency. 2002. Summary of Biologic Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington DC.

U.S. Environmental Protection Agency. 2007. Wadeable streams assessment: A collaborative survey of the Nation's streams. EPA 841-B-06-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

U.S. Environmental Protection Agency. 2009. An Urgent Call to Action: Report on the State-EPA Nutrient Innovation Task Group. U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2009. National Rivers and Streams Assessment: Field Operations Manual. EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2010. Using stressor-response relationships to derive numeric nutrient criteria. Rep., U. S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2011. Working in partnership with states to address nitrogen and phosphorus pollution through use of a framework for state nutrient reductions. A memorandum from N. K. Stoner Acting Assistant Administrator, US Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2012. Water Quality Standards Handbook: Second Edition. EPA-823-B-12-002. U.S. Environmental Protection Agency, Washington DC.

U.S. Environmental Protection Agency. 2014. Expert Workshop: Nutrient Indicators in Streams: Proceedings April 16–18, 2013. EPA-822-R-14-004. United States Environmental Protection Agency, Washington, DC.

Uehlinger, U., B. Kawecka, and C. T. Robinson. 2003. Effects of experimental floods on periphyton and stream metabolism below a high dam in the Swiss Alps (River Spol). Aquatic Sciences 65(3):199–209.

Underwood, A. J. 1989. The analysis of stress in natural populations. Biological Journal of the Linnean Society 37: 51-78.

Underwood A. J. 1996. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Pages 151-175 *In*: Schmitt R. J., and C. W. Osenberg (editors) Detecting ecological impacts: concepts and applications in coastal habitats. Academic Press, San Diego, USA.

United Nations Environment Programme (UNEP). 1992. Rio Declaration on Environment and Development. (http://www.unep.org/Documents.multilingual/Default.asp?DocumentID=78&ArticleID=1163)

Valderrama, J. C. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. Marine Chemistry 10(2):109–122.

Vallentyne, J. R. 1974. The algal bowl—lakes and man. Ottawa, Miscellaneous Special Publication 22, Department of the Environment.

Van de Bogert, M. C., S. R. Carpenter, J. J. Cole., and M. L. Pace. 2007. Assessing pelagic and benthic metabolism using open channel measurements. Limnology and Oceanography: Methods 5:145–155.

Vanote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130–137.

Van Sickle, J., and S. G. Paulsen. 2008. Assessing the attributable risks, relative risks and regional extents of aquatic stressors. Journal of the North American Benthological Society 27(4):920–931.

Van Sickle, J., J. L. Stoddard, S. G. Paulsen, and A. R. Olsen. 2006. Using relative risk to compare the effects of aquatic stressors at a regional scale. Environmental Management 38:1020–1030.

Verheyen D., N. Van Gaelen, B. Ronchi, O. Batelaan, E. Struyf, G. Govers, Merckx R, Diels J. Dissolved phosphorus transport from soil to surface water in catchments with different land use. Ambio. 2015 Mar 1;44(2):228-40.

Vieira, N. K. M., N. L. Poff, D. M. Carlisle, S. R. Moulton II, M. K. Koski, and B. C. Kondratieff. 2006. A database of lotic invertebrate traits for North America: U.S. Geological Survey Data Series 187. Available at: http://pubs.water.usgs.gov/ds187.

Vollenweider, R. D. 1968. The scientific basis for eutrophication of lakes and streams, with particular reference to phosphorus and nitrogen as eutrophication factors. Technical Report OECD. Paris. DAS/CSI/68, 27:1182.

Wagenhoff, A., C. R. Townsend, and C. D. Matthaei. 2012. Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. Journal of Applied Ecology 49(4):892–902.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss. 2003. Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4(1):5–17.

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

Walther, D. A., and M. R. Whiles. 2011. Secondary production in a southern Illinois headwater stream: relationships between organic matter standing stocks and macroinvertebrate productivity. Journal of the North American Benthological Society 30(2):357–373.

Wang, L., D. M. Robertson, and P. J. Garrison. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. Environmental Management 39:194–212.

Webster, J. R., S. W. Golladay, E. F. Benefield, D. J. D'Amgelo, and G. T. Peters. 1990. Effects of forest disturbance on particulate organic matter budgets of small streams. Journal of the North American Benthological Society 9:120–140.

Webster, J. R., and J. L. Meyer. 1997. Organic matter budgets for streams: a synthesis. Journal of the North American Benthological Society 16:141–161.

Webster, J. R., J. B. Waide, and B. C. Patten. 1975. Nutrient cycling and the stability of ecosystems. Pages 1–27 *in* F. G. Howell, J. B. Gentry, and M. H. Smith (editors). Mineral Cycling in Southeastern Ecosystems. ERDA CONF-740513. Reprinted as pages 136–162 *in* H. H. Shugart and R. V. O'Neill (editors). Systems Ecology. Benchmark Papers in Ecology. Dowden, Hutchinson, and Ross Inc., Stroudsburg, Pennsylvania.

Wetzel, R. G. 1975. *Limnology*. W. B. Sanders Publishing, Philadelphia, USA.

Wikle, C. K., and J. A. Royle. 1999. Space-time models and dynamic design of environmental monitoring networks. Journal of Agricultural, Biological, and Environmental Statistics 4:489–507.

Wikle, C. K., and J. A. Royle. 2005. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. Environmetrics 16:507–522.

Wright, J.F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Australian Journal of Ecology*2 0*(1): 181-197.

Yamamuro, A. M., and G. A. Lamberti. 2007. Influence of organic matter on invertebrate colonization of sand substrata in a northern Michigan stream. Journal of the North American Benthological Society 26(2):244–252.

Ylla, I., A. M. Romani, and S. Sabater. 2012. Labile and recalcitrant organic matter utilization by river biofilm under increasing water temperature. Microbial Ecology 64(3):593–604.

Young, R. G., and A. D. Huryn. 1999. Effects of land use on stream metabolism and organic matter turnover. Ecological Applications 9:1359–1376.

Young, R. G., C. D. Matthaei, and C. R. Townsend. 2008. Organic matter breakdown and ecosystem metablism: functional indicators for assessing river ecosystem health. Journal of the North American Benthological Society 27(3):605–625.

Young, R., C. Townsend, and C. Matthaei. 2004. Functional indicators of river ecosystem health. Cawthorn Institute. http://www.cawthron.org.nz/media_new/publications/pdf/2014_07/FunctionalIndicatorsGuideFinal_2.pdf.

Ziliak, S. T., and D. N. McCloskey. 2009. The cult of statistical significance. Pages 2302–2316 *in* Proceedings of the Joint Statistical Meetings, Section on Statistical Education, 3 Aug 2009, Washington, DC.

Zimmerman, D. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17:635–662.

SECTION 4

Comments: Technical Review Team

Placeholder

[Note: David send comments already, but I need to get a Word version in order to be able to incorporate. There is a copy of the .pdf in the Support Documents folder in the P drive.]

SECTION 5

Public Comments

Placeholder

SECTION 6

Appendices

Appendix A: Responses to EPA Comments

Division of Water Quality Responses to EPA Comments (May 1, 2019)

The United States Environmental Protection Agency (EPA) provided technical comments to Utah's Division of Water Quality (UDWQ) about the proposed headwater numeric nutrient criteria (NNC) and its underlying rationale. EPA did not raise any concerns on the NNC elements intended to protect recreational uses (UAC R317-2-14.7), so the responses to comments focus on those NNC elements intended to be protective of aquatic life uses (UAC R317-2-14.8).

EPA raised several concerns about the ecological relevance of the statistically derived response thresholds and their ability to ensure protection of aquatic life uses in headwater streams. UDWQ addresses specific concerns by expanding on the information provided in the supporting documentation in the Detailed Responses below.

EPA also makes several recommendations which involve requests for additional clarification in supporting materials for the proposed headwater numeric nutrient criteria (NNC). Additional EPA recommendations include providing additional supporting materials such as the Standard Operating Procedures (SOPs) for filamentous algae cover data collection and the approach that UDWQ intends to follow for criteria implementation. While these materials have already undergone extensive technical review, UDWQ appreciates the fresh perspectives provided by stakeholders outside of these discussions. EPA is particularly well suited to highlight circumstances where additional details would provide additional rationale that the proposed criteria align with regulatory water quality standard requirements (40 C.F.R. Part 131). UDWQ has revised the supporting documentation—*Technical Support Document: Utah's Nutrient Strategy* (hereafter TSD) and *Proposed Nutrient Criteria: Utah's Headwater Streams* (hereafter Proposal)—to further clarify the underlying rationale behind the NNC. Additional requested documents will be submitted in conjunction with the rulemaking packet provided to EPA for formal review and approval in accordance with section 303(c) of the Clean Water Act (CWA).

Inclusion of the recommended clarifications in supporting documentation will improve the strength of the underlying rationale. However, the comments do not substantively alter the proposed rules or supporting documentation, nor do they raise concerns about the scientific defensibility of the rules as an important regulatory tool that will help the agency maintain and protect water quality in Utah's headwater streams. The proposed headwater NNC does not remove any existing protections from adverse effects caused by excessive nutrient enrichment. Instead, they establish additional protections that can be used to identify adverse effects of nutrient enrichment. The 2015 confirmation investigation (TSD, Chapter 13) confirmed that the NNC were able to identify headwater streams exhibiting adverse effects of nutrient enrichment.

Many headwater streams have naturally-occurring physical attributes that provide natural protections form the effects of nutrient enrichment and this sometimes obscures otherwise significant stressor-response (S-R) relationships. However, the inability of the NNC to identify adverse effects in these circumstances is not evidence that the criteria lack protections for aquatic life uses; in fact, these observations are evidence that combining nutrients with ecological responses is needed to avoid an unreasonable number of false-positive impairment determinations. The responses included in the NNC facilitate an evaluation of the two most important causal pathways leading to degradation of aquatic life uses. Among all of the responses that UDWQ evaluated in the TSD, these responses were also most easily integrated into routine monitoring and assessment programs.

Detailed Responses

The following section provides detailed responses from UDWQ to comments on the NNC submitted by EPA. EPA comments are provided in *blue, italic text*. Responses from UDWQ are in the black text that follows. In some cases, the sections provided by the commenter are subdivided (capital letters) to facilitate responses to major subtopics. For each section, EPA made specific recommendations to UDWQ, which is followed by actions taken by UDWQ, if any, to fulfill each request.

1. Derivation of metabolic thresholds

1A

The EPA has been unable to replicate the analysis used by UDWQ to establish thresholds for GPP and ER as the information provided does not fully describe how the stream metabolism thresholds were derived. The approach used to derive any water quality criterion or threshold, and their supporting data and methods should be transparent and reproducible.

UDWQ agrees that it is critically important for all analyses to be transparent and reproducible. For these reasons all data were made available to interested parties when draft of support documents were released for comment. The methods in the Stream Metabolism chapter of the TSD have been clarified. UDWQ also recalculated the thresholds to ensure that the thresholds could be replicated. This analysis, along with the underlying data and analytical code, will be provided to EPA—along with other supporting materials—when the NNC are submitted to EPA for approval.

1B

Chapter 5 (Pages 54-56, TSD) indicates that the proposed GPP and ER thresholds were derived using nonparametric deviance reduction. The rationale for using deviance reduction methods is not clear. First, Figure 5.2 shows a linear relationship among nutrients, GPP, and ER without any obvious change points in the nutrient-metabolism relationship. Given this linear relationship, it is unclear why UDWQ used a change point analysis rather than linear interpolation methods to derive thresholds.

As explained in the TSD, TP and TN concentrations below the lowest TP and TN NNC thresholds are indistinguishable from reference conditions. Ecological significance is based on the assumption that reference conditions are fully protective. This assumption is also the basis of EPA's (2000) distribution approaches (see *Frequency Distribution Approaches for Setting Nutrient Criteria* in the TSD). Ecological significance was also explicitly considered for the other thresholds. The cited quote from the TSD is accurate but the statement precedes a discussion about the various evaluations that UDWQ conducted to establish the ecological relevance of the statistical thresholds. UDWQ evaluated ecological significance by comparing the thresholds with related water quality standards and assessment indices. In the case of the metabolism threshold, this was done by comparing ER thresholds with Utah's DO criteria. In addition, UDWQ compared the proposed thresholds against independent investigations reported in the scientific literature.

One way to establish the ecological relevance of statistically-derived groups is to evaluate whether or not the other groups correspond with independently derived measures of stream health in a manner consistent with ecological theory. One problem often attributed to eutrophication is low levels of DO. This occurs because additional nutrients increase the abundance or biomass of autotrophs (plants, algae) and heterotrophs (microbes, fungus). The resulting increase in carbon, coupled with additional macronutrients (N, P) fuels increased decomposition. These macronutrients also increase the abundance of microbes and other organisms that decompose these carbon sources. These organisms all respire, consuming oxygen in the process, which can result in DO concentrations that are low enough to be harmful to stream biota. These low DO problems are often particularly apparent during the nighttime when DO consumption is not offset by oxygen produced via primary production. ER is a direct measure of the oxygen consumption associated with the consumption of the additional carbon made available via higher GPP rates. Utah's DO criteria were derived independently, based on oxygen levels found to be harmful to aquatic life, as opposed to underlying processes (GPP, ER) that cause low DO conditions in streams. However, given the cycle discussed above, it follows that the number of excursions below DO water quality standards should increase predictably from low to high ER groups and this is exactly what was documented in the TSD (pp. 57-60). Importantly, no excursion below the 30-day DO criterion occurred among any of the streams in the low GPP and ER groups that define the ecological boundaries of the headwater NNC. The lack of DO violations in this group, but not higher groups, suggests that the lower thresholds are real and values below those thresholds are protective of aquatic life uses.

Another way that UDWQ established ecological relevance of the stream metabolism groups is by comparing the results presented in the TSD against those reported in the scientific literature. These comparisons are presented in the Discussion section of the metabolism chapter in the TSD (pp. 61-64) and also in the proposal. The GPP and ER thresholds in the proposed criteria align with those proposed by other investigators. As discussed in the TSD, these thresholds are also reflective of the first change point, which UDWQ interprets as the point where background GPP and ER rates can be distinguished from naturally-occurring conditions.

EPA Recommendations and UDWQ Responses

To provide additional clarification and to document the scientific defensibility of the criterion, the EPA recommends revision of the TSD language to more accurately and clearly explain the analyses completed including:

- *Providing a step-wise description of the analytical methods used to derive the GPP and ER thresholds. Please provide sufficient detail to ensure the analytical methods can be replicated and the threshold selection process is transparent.*

UDWQ reviewed early TSD drafts to better understand how the TN, TP and metabolism thresholds were derived. The methods and results of Chapter 5 in the TSD have been revised. UDWQ also recalculated the thresholds to ensure that the thresholds could be replicated. This analysis, along with the underlying data and analytical code, will be provided to EPA—along with other supporting materials—when the NNC are submitted to EPA for approval.

- *Documenting how the statistical properties of deviance reduction analysis support the following statement: "Two thresholds were identified; these were expected to provide ecologically meaningful information, with the first threshold corresponding to a departure from the range of natural (reference) conditions, and a second, higher threshold representing an appreciable alteration to GPP or ER processes." (Page 51, TSD).*

As explained in the TSD, TP and TN concentrations below the lowest TP and TN NNC thresholds are indistinguishable from reference conditions. Statistically, the threshold is reflective of the first significant change point in the relationship between ambient nutrient concentrations and stream metabolic rates. UDWQ assumed that GPP and ER rates below this threshold were reflective of naturally occurring conditions. The assumption that the lowest rates are reflective of natural conditions is similar to *post hoc* reference site selection based on water chemistry, which has been used extensively by EPA and state agencies. Ecological significance is based on the assumption that reference conditions are fully protective. In addition, sites below and above the ER threshold were evaluated against Utah's numeric DO criteria (Proposal, Figure 8) which validated that the lower threshold is protective of potentially stressful DO conditions. The assumption that reference conditions are reflective of natural conditions is also the basis of EPA's (2000) distribution approaches, and a similar analysis conducted by UDWQ that confirm that the nutrient concentrations below the TN and TP concentrations below the lower metabolic thresholds fall within the range of reference conditions. The higher threshold represents the next statistically valid change point in the relationship between nutrient concentrations and stream metabolism, which seemed to distinguish streams with atypically high GPP or ER. The upper thresholds assisted with data interpretation and were otherwise not integral to the NNC.

- *Describing whether UDWQ explored the use of linear interpolation methods (either from the literature, reference, or the relationship of these metabolic indicators with other endpoints) to derive the GPP and ER thresholds.*

UDWQ provided additional details with respect to the underlying rationale behind the analytical methods used to derive the GPP and ER in the responses to section 2C above. The linear relationship among nutrients and metabolism responses was only apparent after natural log transformations of both nutrients and responses. Hence, these relationships are exponential, not linear. While statistically significant, there was also considerable scatter due to error and the influence of covariates. UDWQ used the linear regression in an exploratory fashion to efficiently confirm that an underlying relationship between the variables could be measured. Once this relationship was confirmed, UDWQ then applied more sophisticated analyses to characterize the relationship and to specifically identify change points.

With respect to additional responses, UDWQ explored the relationship between autochthonous organic matter standing stocks and metabolic rates (TSD, Chapter 6). The relationship between ambient nutrient concentrations and structural responses was not explored in the same way because UDWQ explored these relationships using existing data sources to maximize the sample size and generate the most robust structural response relationships possible. The results of all of these analyses are presented in the TSD (Chapters 2-7).

- *Explaining the relationship of the statistically-derived nutrient groups identified in Table 5.2 (TSD, page 55) to the nutrient thresholds proposed as criteria in Table 12.1 (TSD, page 139).*

Table 5.2 provides the nutrient concentrations that, on average, most strongly distinguished among streams with low, moderate and high metabolic rates. These thresholds were among many that UDWQ used as multiple lines of evidence (see Proposal and TSD, Chapter 12) to inform the nutrient thresholds. UDWQ identified lower nutrient thresholds that were reflective of the upper range of background conditions and then collectively evaluated the nutrient thresholds from all responses to ensure that these lower concentrations would be protective of several different indicators of aquatic life use support. The nutrient thresholds associated with changes in metabolic responses were among the indicators evaluated in this review of the proposed criteria. UDWQ edited the TSD to provide a more thorough explanation of these relationships and discusses the ecological significance of these thresholds in the responses to section 2B of this comment letter.

- *Explaining how the GPP and ER criteria thresholds relate to distinct changes in the aquatic life community, structure or function.*

The TSD discusses linkages among metabolic responses and nutrient enrichment extensively, but UDWQ has provided additional information on the linkages among metabolic responses and aquatic life uses in the responses to comments. In some respects, linkages to ecosystem functions are self-evident because GPP and ER are considered fundamental ecosystem functions. Chapter 2 of the TSD discusses linkages between these indicators, the protection of aquatic life uses, and Utah's water quality standards. Chapter 5 provides additional details on the ecological importance of GPP and ER responses with further elaboration provided in the responses to comment 2A. Linkages between stream metabolic rates and higher trophic levels are intrinsically indirect, but virtually all nutrient-related adverse impacts to the composition of stream assemblages are initiated by increases in autotrophic (GPP) or heterotrophic (ER) production. Hence, many of the nutrient-related alterations to the composition of macroinvertebrates and diatom assemblages documented by UDWQ (TSD, Chapter 7) are indirectly attributable to increases in GPP or ER.

With respect to the specific GPP and ER thresholds, several lines of evidence can be used to establish their ecological relevance. The metabolic thresholds are reflective of the lowest statistically valid change-points associated with ambient nutrient concentrations. UDWQ interpreted this response as an inflection point that distinguishes background GPP and ER rates from those altered by nutrient enrichment. The corresponding TN and TP concentrations are roughly equivalent of the 75th percentile of reference site distributions (TSD, Chapter 11), which provides additional support for the interpretation with respect to the NNC stressor. Biological integrity cannot be measured directly because it is a construct intended to capture many different facets of stream condition. Instead, resource managers have relied on indicators reflective of important aspects of stream condition. After calculating ER thresholds, UDWQ compared ER rates with Utah's DO criteria and found that excursions below the criteria were much more likely above the ER NNC threshold (TSD, Chapter 5). UDWQ also found a significant relationship between nutrients (TN and TP) and standing stocks of autochthonous organic matter (OM). Streams with low levels of autochthonous organic matter were much more likely to show excursions below DO criteria and almost always exhibited ER rates below the proposed ER threshold (TSD, Figure 6.3, panel D). In contrast, the vast majority of streams with high autochthonous OM biomass had ER rates greater than the proposed NNC threshold.

Many studies have demonstrated linkages between stream metabolic rates and various stressors, including nutrients. However, few translate their results into ER or GPP condition classes. Young et al. (2006) suggested that GPP (4 g O_2/m^2/day) and ER values (5.5 g O_2/m^2/day) differentiated streams in good vs. fair condition. These recommendations are quite similar to the lower thresholds derived from Utah streams (GPP of 6 and ER of 5 g O_2/m^2/day). Another study (Izagirre et al. 2008), compared the thresholds proposed by Young to metabolic values obtained from 19 streams in Spain and suggested that the proposed thresholds may be overly conservative because they were exceeded in all of their most oligotrophic and biologically healthy streams (average GPP of 9 g O_2/m^2/day). There are many other causal pathways that could potentially link alteration of GPP and ER rates to degradation of uses (see responses to 3A above). UDWQ was unable to evaluate the ecological relevance of GPP and ER thresholds for every causal pathway, but those that were evaluated demonstrate that the proposed thresholds are appropriately protective. In contrast, data have not been presented that provide evidence that lower rates are needed to be protective of biological uses.

2. *Issues related to calculation of stream metabolism metrics or interpretation of results*

2A

Utah's TSD states that calculation of metabolic rates should be routine using typically collected data:

"DWQ's monitoring and assessment program has practical constraints that were considered when developing the proposed NNC, which were defined to facilitate routine data collection of the parameters necessary for making decisions about support of aquatic life in headwater streams. This means the protection of aquatic life uses will not be impeded by a lack of sufficient data or information necessary to make regulatory decisions." (Page 24, TSD) [emphasis added]

"In addition, there are rare circumstances in which metabolism models cannot accurately calculate GPP and ecosystem respiration (ER), because diel changes in dissolved oxygen (DO) are insufficient for making reliable reaeration estimates." (Page 152, TSD) [emphasis added]

UDWQ's comment about the circumstances where metabolism models could not be generated (or the inability to generate accurate models) should have specified that this is only true in circumstances where minimal GPP and ER do not constitute a threat to aquatic life uses. Circumstances where metabolism models cannot be easily generated are caused by an inability to accurately model an estimate of atmospheric reaeration, due to relatively small diel variations in DO. This typically occurs in streams that are very turbulent, highly unproductive, or have very high rates of atmospheric reaeration. In all such circumstances, both ER and GPP are low, and although unquantified, can be concluded to be below the NNC thresholds. Such circumstances are analogous to non-detects below the criterion in chemical analyses when the concentration of a pollutant could not be measured because concentrations are below the analytical detection limit.

Once equipment has been purchased, there is nothing that constrains UDWQ's ability to collect the data necessary to model stream metabolism.

As discussed in more detail in responses to Section 3 of the comment letter, the relatively large number of sites where GPP and ER could not be calculated in the 2015 study is a reflection of the fact that many of these streams have attributes that make them naturally protected from <u>all</u> adverse effects resulting from nutrient enrichment. For example, high channel shading decreases the likelihood of enrichment altering the structure of stream autotrophs and subsequent modification to higher trophic levels in stream food webs or increases in GPP rates. High channel gradients decrease the residence time of nutrients, similarly minimizing enrichment conditions that could alter stream autotrophs. Higher channel gradients also increase reaeration, minimizing diel DO flux in the process. Similar analogies could be made for other covariates. If streams have physical attributes that make them naturally protected from the adverse effects typically caused by nutrient enrichment, the inability to easily measure GPP and ER does not diminish the protectiveness of the headwater criteria; it is simply reflective of a lack of local adverse effects and a demonstration of the value of combined criteria. UDWQ does not believe that it is appropriate to presuppose impaired conditions at moderately enriched streams with naturally-occurring attributes that are protective of adverse effects resulting from nutrient enrichment. The proposed criteria also include upper TN and TP thresholds that do not require responses, so evaluations of enrichment can still be made even if metabolism data are unavailable.

If the NNC were adopted, filamentous algae and metabolism metrics would not be the only responses available to UDWQ for purposes of identifying nutrient-related impairments. Other numeric criteria are routinely evaluated that can be used to identify impairments along several causal paths linking nutrient enrichment to degradation of aquatic life uses. For example, Utah's streams are naturally alkaline (pH > 8) and violations of Utah's upper pH criterion of 9 (UAC R317-2-17) sometimes occur when rates of primary production are high due to reductions in water column CO_2. This increase in pH has been associated with adverse effects to fish and also increases the relative toxicity of ammonia. Similarly, Utah has both acute and chronic DO criteria, which are used to identify adverse nutrient-related effects along related causal paths. Both DO and pH are routinely measured during every sampling event for all of Utah's waters. Utah also routinely conducts biological assessments on streams using O/E, and the TSD (pp. 83-104) linked nutrient enrichment to an increased probability of identifying biologically degraded waters using this assessment method. These proposed NNC also include protections for recreational uses and the associated evaluations of biomass accrual are another response that will be used to evaluate potential deleterious effects of nutrient enrichment, albeit for an alternative designated use.

2B

Results from headwater streams sampled in 2015 show that two of the three stream metabolism metrics proposed by UDWQ for their combined criterion could <u>not</u> be calculated at approximately 40% of sites. The inability to calculate stream metabolism metrics at many sites is a practical barrier that limits the utility of these indicators. In these situations, UDWQ must rely *on percent filamentous algae to make attainment decisions. In Section 7 of this comment letter, we describe concerns with the protectiveness of the percent filamentous algae threshold. The lack of a clear relationship between the proposed filamentous algae threshold and aquatic life use protection is particularly concerning for the 40% of headwaters sites where attainment decisions will likely be based on this indicator. Since the response indicators serve a critical function in the moderate range of TN and TP values (essentially off-ramping the use of the TN and TP criteria thresholds), it is critical to establish their use, effectiveness and reliability in headwater streams ecosystems.*

UDWQ is not concerned about the relatively high number of streams where metabolism responses could not be calculated because these “non-detects” are reflective of circumstances where elevated GPP and ER are not reflective of adverse conditions (see response to 3A). In many of these cases where metabolism was not calculated from the 2015 data, it may have been possible to generate the data using more sophisticated models. However, UDWQ did not spend the resources to calculate these results because the objective of that investigation was to validate the NNC thresholds, and in that context, it was clear that these sites would fall below GPP and ER thresholds.

2C

Young proposes that low ER values (<0.8 g $O_2/m^2/d$) indicate poor stream health. Based on this information, please explain whether values <0.8 were considered in either deriving the stream metabolism thresholds using the 2010 dataset or in reviewing the 2015 data.

Very low rates of ER can be reflective of poor stream health because they may indicate widespread extirpation of aquatic life or a loss of hydrological connection between surface waters and the hyporheic zone due to significant and widespread sedimentation. UDWQ has not included low ER rates in the proposed NNC for a couple of reasons. First, UDWQ expects that these hypothetical situations are rare among headwater streams because they are typically reported to occur in the most highly degraded streams. The causes of these deleterious responses are generally not attributable to nutrients, so this would be out of scope with respect to the proposed NNC. Low thresholds for ER could potentially be useful as a general indicator of stream health but is unlikely to be useful for identifying nutrient impairments. Other water quality indicators will be more sensitive to widespread loss of aquatic life, and it is unlikely that a low ER component to the criteria would identify an impairment that would not be identified using other routinely collected water quality parameters. Second, as discussed in responses earlier in this section, reliable measures of very low ER rates can be problematic without a direct measure of reaeration. Directly measuring reaeration in the field is a lengthy process and would require an expansion of laboratory capacity with new analytical methods, which are both practical constraints that limit their routine collection.

EPA Recommendations and UDWQ Responses

The EPA recommends:

- *Reevaluating whether GPP and ER will serve as functional response indicators that can be reliably collected in headwater streams.*

  DWQ has reviewed the efficacy of GPP and ER as functional response indicators and have determined that these indicators can reliably be collected through our routine monitoring programs and utilized as designed in the combined criteria framework.

- *Evaluating whether some of the results presented in the TSD are valid for inclusion.*

  DWQ has reviewed the results in the TSD. Two sites have been flagged with respect to metabolism model results. None of these sites had high nutrients, GPP, or ER and therefore, their exclusion from the analysis does not affect the interpretation of study results with respect to the proposed NNC.

- *Considering the implications of low ER values.*

  UDWQ considered the implications of low ER values and provided a more detailed explanation of whether such responses are appropriate for the NNC in the response to section 2C. With the exception of the unusual circumstance of e.g., overt toxicity or very high levels of anthropogenic stress, low ER indicates that nutrients are not causing adverse effects to stream biota. The potential for such overt toxicity is highly unlikely in headwaters, and if such conditions were identified, the cause would not be nutrient-related, so these responses would be an inappropriate addition to nutrient criteria.

- *Providing a copy of Utah's stream metabolism Standard Operating Procedures (SOPs) to the EPA as part of the submission package.*

  The requested documentation will be submitted in conjunction with the rulemaking packet provided to EPA for formal review and approval in accordance with section 303(c) of the CWA.

3. Responsiveness of stream metabolism indicators to nutrient pollution in the headwaters

3A

The EPA is concerned that the stream metabolism indicators are not sensitive to nutrients. 40 C.F.R. § 131.11(a) requires that "criteria must be based on sound scientific rationale and must contain sufficient parameters to protect the designated use." (Emphasis added.) The EPA considers response indicators that are sensitive to nutrient pollution as examples of the types of "sufficient parameters" needed to protect the designated use, which for Utah's headwater *streams is aquatic life. The EPA's guidance document entitled "Guiding Principles on an Optional Approach for Developing and Implementing a Numeric Nutrient Criterion that Integrates Causal and Response Parameters" recommends that, for a combined criterion, response indicators should be sensitive to nutrients such that a predictable relationship exists between the two. Response indicator thresholds should be linked to nutrient concentrations and demonstrated to protect uses (i.e., aquatic life uses). Without this quantified relationship, response indicator thresholds may be unable to ensure use protection.*

UDWQ divided responses to these comments into several subsections to address major elements of the comments. These responses are likely more detailed than the comment strictly requires because the comment reflects several important themes related to other comments in the letter.

Metabolism Responses are Linked to Nutrient Concentrations

UDWQ demonstrated a predictable relationship among stressors (TN, TP) and all of the responses evaluated, including GPP and ER (see Figure 5.2 in Section 5A of these comments).

Metabolism Responses are Heuristically Linked to Aquatic Life Uses

The central tenet of combined criteria is the simultaneous consideration of nutrient concentrations and <u>all</u> associated responses. Utah's headwater criteria include upper thresholds that do not require responses to assess, which means that the responses are only relevant to moderately enriched streams. Sensitivity (capability of protecting designated uses) should be evaluated in the context of false positive and false negative decisions with false negatives being of greater concern. UDWQ concludes that false negative decisions in the moderate range of nutrient enrichment are unlikely. The headwater NNC includes filamentous algae cover, GPP and ER responses as explicit protections of aquatic life uses and excursion of any one of responses is interpreted as an excursion. Additional implicit responses related to nutrient enrichment responses are also routinely evaluated against independent numeric criteria including DO and pH (which are already evaluated with existing numeric criteria) and biological assessments based O/E macroinvertebrates scores (see also response to Section 2 comments). Consideration was given to adding implicit responses to the proposed NNC, but UDWQ determined—in consultation with regional EPA staff—that retaining their independent evaluation would be <u>more protective</u> of designated uses. If any of these explicit or implicit responses indicate degradation at a moderately enriched stream, it would be considered impaired. It is UDWQ's position that assessments based on the combination of these responses, combined with the upper nutrient thresholds, are sufficient to identify nutrient-related adverse effects in moderately enriched headwater streams. This was confirmed by the 2015 confirmation investigation (TSD, pp. 152-166) where the NNC successfully identified a considerable number of nutrient-related impairments. The combination of nutrients and responses in the NNC is independently protective. The <u>additional</u> responses incorporated into the NNC provide additional lines of evidence that UDWQ can use to identify nutrient-related problems. EPA has provided no evidence in the comment to support the conclusions that NNC "<u>may</u> be unable to support use protection."

Aquatic life uses in Utah and elsewhere were established to help ensure the protection and maintenance of the biological integrity of waters as mandated by the Clean Water Act. Biological integrity is an integrative construct used to convey a multifaceted concept of ecosystem health. Ecologists often partition biological integrity into structural and functional components, which is reflected in the most widely accepted definition (emphasis added):

> "*The ability to support and maintain a balanced, integrated adaptive assemblage of organisms having species composition, diversity, and functional organization comparable to that of natural habitat of the region*" (Karr and Dudley 1981, Karr et al. 1986).

Structural integrity reflects elements of biological degradation that alter the composition and relative abundance of stream biota, whereas functional integrity elements reflect alterations to elements integral to important ecosystem functions or processes. Ecosystem processes describe the flow of energy through food webs, which is generally measured through the production or consumption of carbon. GPP and ER are fundamental measurements of these important processes.

GPP and ER are linked to aquatic life uses because they are integral to the construct of biological integrity. This tie to aquatic life use support is an implicit tenet of all state biological assessment programs. For example, Utah assesses aquatic life use support using O/E, which quantifies local extinction of macroinvertebrate taxa resulting from human-caused stressors, other states use similar metrics. EPA has not required Utah, nor other states, to demonstrate that the loss of invertebrate species are "necessary aquatic organisms" because their loss, in and of itself, is considered to diminish central aspects of biological integrity. GPP and ER are similarly linked to aquatic life uses as fundamental mechanistic measures of stream ecosystem functions.

EPA (2010) guidance emphasizes the importance of conceptual models when using stressor-response models to establish NNC. One purpose of these models is to demonstrate the specific causal pathways that link nutrient enrichment to degradation of aquatic life uses (e.g., TSD, Figure 2.1). Linkages between nutrient enrichment and degradation of uses are often complex, often involving several interim steps before deleterious effects to uses occur, and this sometimes involves several different, potentially interrelated, pathways. One way that these models can be used is to visualize how ecological responses—in this case GPP and ER—are integral to the protection of uses. Another way that these models can be used is to identify responses that are as proximate to nutrient enrichment as possible, because these responses are assumed to generate more accurate and sensitive S-R relationships.

GPP and ER are direct measures of fundamentally important ecological alterations to nutrient enrichment and are as proximate to nutrient enrichment as possible. Almost all causal pathways linking nutrient enrichment to aquatic life uses start with increases in the abundance of autotrophs or heterotrophs. Many causal pathways related to autotrophs are caused by increases to primary production rates, which GPP quantifies.

Similarly, many heterotrophic responses are caused by increased microbial (e.g., bacteria, fungus) biomass, and while decomposition is not the only a function of these organisms, it is among the most important with respect to the health of stream food webs. ER directly measures rates of organic matter processing, which is particularly important to headwater streams since they are typically net heterotrophic due to most of the energy needed to support life comes from external (allochthonous) sources. Also, excessive ER depletes DO, degrading aquatic life uses. To suggest that GPP and ER are not linked to aquatic life uses ignores the fact that these indicators measure processes that are directly linked to all of the nutrient related causal pathways with the potential to degrade aquatic life uses. Moreover, given that nearly all adverse effects are initially caused by increases in plant and microbial growth, other deleterious effects are unlikely to occur without adverse effects to one of these proximate responses.

Multiple Lines of Evidence Suggests that the GPP and ER Thresholds are Appropriately Protective

Ambient nutrient concentrations for sites below the lower metabolic thresholds are roughly equivalent of the 75th percentile of reference site distributions (TSD, Chapter 11) and UDWQ considers such conditions to be reflective of full support of aquatic life uses (see Section 1 responses for additional details). Reference conditions are integral to the demonstration that NNC derived from frequency distribution approaches are protective, but the approaches are limited with respect to the overarching CWA objective to protect biological integrity. Unfortunately, biological integrity cannot be measured directly because it is a construct that is intended to capture many different facets of stream condition. Instead, resource managers have relied on indicators reflective of important aspects of stream condition. In the case of GPP and ER, the relevance of the NNC thresholds as indicators of important biological integrity components was evaluated in a couple of ways. After calculating ER thresholds, UDWQ compared ER rates with Utah's DO criteria and found that excursions below the DO criteria were much more likely above the ER NNC threshold (TSD, Chapter 5). UDWQ also found a significant relationship between nutrients (TN and TP) and standing stocks of autochthonous organic matter (OM). Streams with low levels of autochthonous organic matter were much less likely to show excursions below DO criteria, and almost always exhibited ER rates below the proposed ER threshold (TSD, Figure 6.3, panel D). In contrast, the vast majority of streams with high autochthonous OM biomass had ER rates greater than proposed NNC threshold. UDWQ was unable to evaluate the ecological relevance GPP and ER thresholds for every causal pathway, but those that were evaluated support the proposed thresholds as appropriately protective. In contrast, evidence has not been presented from EPA or others that suggest that lower rates are needed to be protective of biological uses.

NNC Derivation Involved an Evaluation of Most Candidate Responses

In 2013, EPA convened a symposium of international experts with the objective of establishing the most appropriate and sensitive responses of nutrient enrichment in streams (EPA 2014). UDWQ reviewed the findings of this symposium and almost all of the responses recommended by these experts were evaluated in the process of establishing the proposed headwater NNC; it remains unclear what additional responses should be added. The only potential responses that UDWQ did not extensively evaluate were algal and nutrient-specific macroinvertebrate metrics, and UDWQ is willing to add these as additional responses to the NNC if empirically-supported thresholds are identified in the future. In the interim, UDWQ maintains that the addition of these NNC to Utah's water quality standards are scientifically sound for protecting waters from the effects of nutrient enrichment, and therefore protective of the support of aquatic life and recreational uses.

3B

UDWQ examined whether the proposed metabolic indicators responded to nutrients or other potential covariates using Random Forest. Table 5.1 (page 53, TSD) summarizes the results of the Random Forest analysis. "Variables with a larger increase in the mean squared error (MSE) are more important determinants of metabolic rates relative to others." (See page 53, TSD). For GPP, the primary predictors of metabolic rates were channel shading (MSE=104.3), slope (MSE=103.1), and basin slope (MSE=75.3) followed by nutrients (MSE = 72.6 for TN and 73.9 for TP). For ER, channel shading (MSE=69.3), total nitrogen (MSE=69.0), and slope (MSE=63.4) were the most important factors influencing ER. UDWQ notes that they further explored the influence of relevant covariates (stream slope, shading, TN, and TP) for both response indicators to improve the accuracy of the metabolism results:

"NDR [nonparametric deviance reduction] revealed significant thresholds at ~1% slope for both ER and GPP. GPP and ER thresholds were also found for percent channel shading, where streams with channel shading less than ~11% had greater mean daily rates of GPP (9.3 ± 5.6 to 3.99 ± 4.1) and ER (8.10 ± 5.5 to 4.31 ± 4.1)." (Page 60, TSD)

These results show that at sampling locations with greater than 1% slope and greater than 11% channel shading these factors have a greater influence than nutrient concentrations on metabolic rates. Information presented in the TSD documented that the majority of the headwater streams sampled in 2015 have slopes that exceed 1% and shading greater than 11%. Therefore, one would expect that the ER and GPP response indicators would have limited utility in signaling nutrient impacts in these headwaters:

"Canopy cover (shading of the stream channel by riparian vegetation) averaged about 50% but ranged from 3% to 87%. Sites were generally of high gradient, averaging about 5% slope, with a range from < 0.5% to just over 20% among all streams." (page 157, TSD)

The TSD documents the limitations in the relationship between the proposed metabolism indicators and nutrients concentrations:

"In general, neither GPP nor ER was high in comparison with the broader investigation used to generate S-R relationships." (Page, 158, Chapter 5, TSD). [emphasis added]

"When GPP and ER rates were compared across the range of nutrients that occur throughout Utah, GPP and ER were found to be relatively robust indicators of enrichment, particularly when channel shading and slope were accounted for (Chapter 5). However, in the lower ranges of nutrient concentrations in headwater streams, these responses were not particularly sensitive measures of enrichment. This follow-up investigation did not document a single stream with elevated GPP, and one stream with elevated ER was identified." (Page 163, Chapter 12, TSD). [emphasis added]

Because of the physical characteristics of headwater streams - cooler, lower nutrient concentrations, greater reaeration (due to steeper slopes), and shadier (less GPP) - it's not surprising that the proposed stream metabolism indicators do not respond to increased nutrient concentrations in headwater streams and the data presented do not provide sufficient evidence of a relationship. Information presented in the TSD confirms that, at the 60% of headwater sites where metabolic rates could be calculated, GPP and ER were not sensitive to nutrient pollution. The EPA is concerned that, if GPP and ER are not sensitive to nutrients, use of these indicators and/or the derived thresholds will not serve to confirm nutrient impairments in these streams (see Figure 5.2 on page 9 below).

A response only needs to be sensitive to enrichment under conditions where a stream is naturally susceptible to the adverse effects that it quantifies. If no response provides evidence of adverse effects, then it is reasonable to conclude that the condition of aquatic life uses in the moderately enriched stream being assessed has not been degraded. The alternative interpretation—implied in this comment: "*will not serve to confirm nutrient impairments in these streams* (emphasis added)"—is to presuppose impairments exist in circumstances where adverse effects have not been demonstrated.

The Random Forest results and related analyses provide additional support for the linkage between nutrient enrichment and stream metabolism. Both TN and TP were among the top 5 explanatory variables for both ER and GPP. The fact that physical characteristics also controlled the magnitude of metabolic responses is not surprising, nor is it problematic from the perspective of ensuring that the NNC are protective of aquatic life uses. Headwater streams that are relatively sensitive to nutrient enrichment (e.g., less channel shading, lower channel gradient) showed an increase in GPP and ER with increasing nutrients. If conditions are amenable to adverse effects and streams are enriched, the potential exists for adverse effects from increased rates of GPP and ER. As discussed in the responses to Section 2 of these comments, the same attributes that make some streams naturally resistant to increases in GPP and ER are also protective of other nutrient responses, i.e., the stream is not impaired. When a stream is moderately enriched and lacks the physical attributes to sufficiently protect against adverse effects, it is possible that other related numeric criteria (e.g., pH, DO) or indicators (e.g., O/E, chl-*a*, filamentous algae cover) would identify nutrient-related problems. If none of these responses indicate adverse effects, then it is reasonable to conclude that the moderately enriched streams had naturally protective attributes and local threats to aquatic life are minimal.

EPA Recommendations and UDWQ Responses

Based on this information, the EPA recommends:

- *Specifying how UDWQ accounted for the influence of channel slope and shading when establishing the GPP and ER thresholds in headwaters.*

The influence of covariates is among the principal reasons for the incorporation of both nutrient and responses in the NNC. Without the incorporation of responses, it would be nearly impossible to identify criteria that were protective of uses in streams where adverse effects are most likely to occur. Approximately half of the moderately enriched streams in the 2015 investigation did not show any indication of adverse effects. Practically speaking, local characteristics like slope and channel shading are difficult to incorporate into regional NNC without compromising their regulatory applicability. The proposed NNC will be applied to steams that do not have nutrient point sources. As a result, the primary regulatory application will be identifying water quality impairments. If a moderately enriched stream has higher slopes or canopy cover this would decrease the likelihood that problems with GPP and ER would manifest. An observation of low GPP and ER in such circumstances demonstrates that these adverse effects do not exist. Conversely, in the physical characteristics of a stream are not naturally protective, the expectation is that thigh rates of GPP and ER would manifest themselves, which is the intent of combining nutrients and ecological responses in the NNC (see also Section 3B responses). As noted in the comments, UDWQ did evaluate the influence of numerous physical characteristics know to control GPP and ER rates and while several were important, TN and TP were also among the top 5 predictor variables. This demonstrates that under the right conditions, nutrient enrichment is an important determinant of these important measures of ecosystem function in streams.

- *Analyzing the 2015 headwaters data to demonstrate that headwater streams are sensitive to nutrient pollution.*

UDWQ did analyze the 2015 data (TSD, Chapter 13) and, as discussed in the responses to comments, these analyses did demonstrate that some headwater streams are sensitive to nutrient enrichment, which the combination of nutrients and responses in the proposed NNC successfully identified. Many of the streams in this study, which focused on streams where previous sampling suggested moderate to high nutrient enrichment, would be considered to be impaired using the proposed NNC. None of these streams would have been identified as impaired using conventional, nutrient-related parameters (e.g. pH, DO). These analyses demonstrated the added protections to aquatic life that these NNC would provide to these important ecosystems.

4. Representativeness and applicability of metabolic indicators from all streams to headwater streams

Sites sampled in 2010 to develop the combined nutrient criterion differ significantly from headwater streams for which the combined criterion are proposed and the TSD does not clearly *demonstrate how these thresholds are protective of headwaters. As discussed above, UDWQ's analysis showed that sampling locations with greater than 1% slope and less than 11% channel shading have a greater influence than nutrient concentrations on metabolic rates. However, UDWQ used data from a statewide study, not headwater streams, to select the proposed response indicators (i.e., GPP and ER), evaluate their sensitivity, and establish thresholds. Streams from the 2010 study were large (average watershed area of 187 mi^2, average discharge of 29 cfs, average width of ~25 feet, average depth of 1.1 feet, average slope of only 1.1%, and average canopy cover of only 27%). (Table 3.2, page 29-30, TSD). Utah conducted additional sampling of headwater sites in 2015; however, data from these events were not considered in the selection of the response indicators or in the derivation of the proposed thresholds. The EPA was not able to locate data summarizing the drainage area, width, depth and slope of these headwater sites to compare to the 2010 study.*

Young (2008) notes: "metabolic rates at potentially impacted sites should be compared with rates measured at (more) pristine sites that are characterized by similar stream order and size" [emphasis added]. If there are significant differences in drainage area and stream orders between sites sampled in 2010 compared to headwater streams, then hydrodynamic differences alone between size classes could confound the ability to use data from the 2010 study to establish a protective combined nutrient criterion for headwaters streams.

Utah's headwaters have fewer enriched sites, particularly at the upper end of ambient nutrient concentrations (i.e., highly enriched streams). Limited responses were expected among moderately enriched headwater streams because they are more likely to have physical attributes that make them naturally protected from adverse effects of nutrient enrichment. In order to avoid missing otherwise ecologically meaningful S-R relationships, it was important to ensure that the derivation of NNC thresholds was made from a dataset with a fairly even number of low, moderate, and highly enriched streams. For example, DO and ER followed a predictable relationship when the underlying data had a sufficient number of streams at the higher end of ambient nutrient concentrations observed in Utah (TSD, Chapter 5), but this relationship was less apparent when data were limited to less enriched and sensitive streams (TSD, Chapter 13). DO deficits are caused by excessive respiration, so the former is likely a more accurate reflection of this S-R relationship. Similarly, a statistically significant relationship between nutrients and stream metabolism (see Figure 5.2 below) becomes obscured when the underlying data are biased toward a higher number of naturally tolerant streams.

EPA Recommendations and UDWQ Responses

The EPA recommends:

- *Explaining in the TSD how the data used relate to the derivation of headwater streams thresholds and ensure aquatic life protection in headwater streams.*

The first 14 chapters of the TSD (~180 pages) contain an extensive explanation of how the data used relates to the preservation of aquatic life and the application of stressor-response relationships to the protection of headwater streams. UDWQ maintains that this documentation, when combined with the Proposal and additional clarifications provided through the responses to these comments, provides sufficient demonstration that the proposed NNC would be a valuable regulatory tool to address the adverse effects of nutrient pollution in headwater streams.

- *Adding a table to the TSD that summarizes the drainage area, width, depth, etc. for the headwater sites sampled in 2015.*

UDWQ made a concerted effort to provide all of the data used to conduct all analyses to all stakeholders, including EPA. The results and interpretation of these analyses have been vetted with stakeholders, including a technical review team. The 2015 investigation was intended as a confirmational investigation, with the stated intent of confirming the proposed NNC. The study was not intended to refine existing thresholds and the study design is not amenable to doing so. Physical habitat characteristics of the 2015 investigation are already summarized in the context of interpreting the data presented and study objectives. If EPA needs additional data and information as part of their formal review of the NNC, UDWQ will provide additional, readily available data and information.

- *Comparing the morphologic and physical characteristics of the 2010 study sites to headwater sites to demonstrate that the streams sampled in 2010 are comparable to headwater streams.*

The TSD already summarizes the physical characteristics of headwater streams collected in the 2015 conformational investigations: “Despite the fact that study sites were constrained to headwater streams, they exhibited a fairly wide range of physical characteristics. While sites were all at relatively high elevations (average of ~2,200 m) relative to all streams in Utah, they spanned a fairly wide range of elevation, from 1,580 to 3,040 m. Canopy cover (shading of the stream channel by riparian vegetation) averaged about 50% but ranged from 3% to 87%. Sites were generally of high gradient, averaging about 5% slope, with a range from < 0.5% to just over 20% among all streams. Given these relatively high gradients, it is not surprising that about 80% of sites had relatively coarse substrates (> 50% of measured cobbles were 64–250 mm or larger).” This breadth of physical conditions highlights the importance that UDWQ placed on selecting S-R study sites that encompassed a range of conditions, particularly those where adverse effects of nutrient enrichment are likely to manifest.

- *Using representative data from the target headwater streams to select response indicators and derive protective thresholds or explain why these data are not relevant.*

UDWQ designed the studies to efficiently determine the ecosystem response thresholds. To establish response indicators based on a e.g., random sample design, was anticipated to require an exceptionally large number of samples because of general lack of significant nutrient inputs to cause measurable responses. Category 1 and 2 waters are Utah's most stringent antidegradation tiers and no new discharges are allowed in Category 1 waters. Compared to downstream waters, these waters have little anthropogenic influence. UDWQs expectation was that measurable nutrient impacts will only occur rarely. Hence, UDWQ noted that the headwater sites selected were biased towards the most impacted sites known. The inability to always quantify ER and GPP in some of the headwater streams confirmed the expected lack of adverse nutrient-related ecosystem responses.

5. *Protectiveness of Proposed Thresholds and Link to Aquatic Life Use Protection*

5A

The TSD does not clearly link key metabolic indicators that Utah has selected for use in the combined criterion to protection of aquatic life uses in headwater streams. The EPA used Figure 5.2 (Page 60, TSD) to further review the nutrient-metabolism relationship in Utah's headwater streams. The EPA superimposed the following information on the figure:

- *Low, medium and high thresholds for TN and TP (orange and red lines)*
- *Mean GPP and ER values for the low, medium, and high nutrient groups (blue arrows)*
- *Thresholds for GPP and ER (black hashed lines)*
- *2015 metabolism results at headwater streams (blue symbols)*

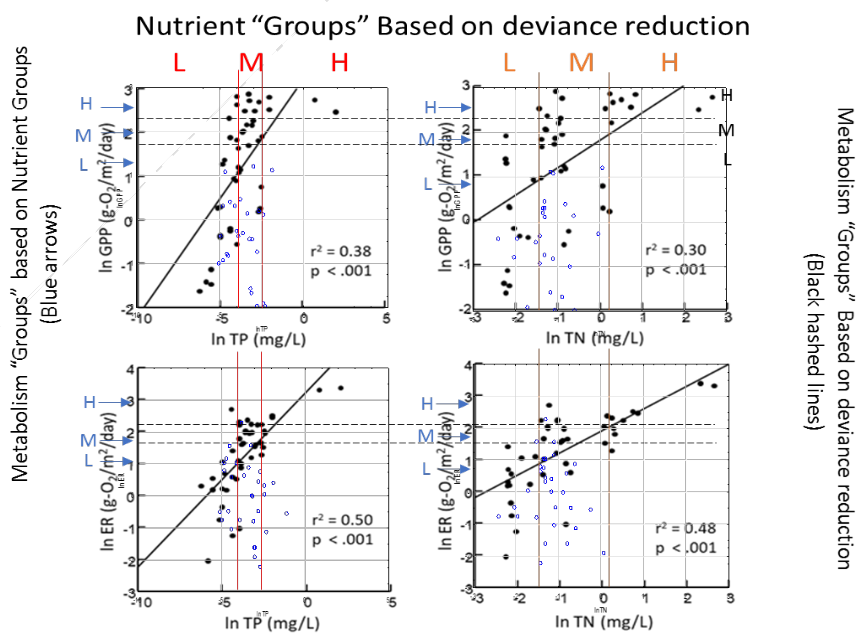


Figure 5.2. Linear relationships between ambient nutrient concentrations (total nitrogen [TN] and total phosphorus [TP]) and gross primary production (GPP, top row) and ecosystem respiration (ER, bottom row).

With this information added to the figure, it is apparent that the proposed ER threshold of 5 g-O_2/m^2/day (Ln value of 1.6) is at or above the upper threshold TP value. Because the upper TP threshold is associated with impaired conditions, it is unclear how the proposed ER threshold demonstrates protection of aquatic life uses.

UDWQ acknowledges that similar to NNC efforts by EPA and other states, infrequent decision errors may occur. With the current state of the science, it is difficult to establish broadly protective N and P criteria when the magnitude of enrichment necessary to elicit adverse effects varies from one stream to the next due to variance in naturally-occurring, local stream attributes. However, UDWQ is unable to duplicate the conclusions of the comment regarding the modified graph.

From the perspective of sensitivity to nutrient enrichment, UDWQ's analysis on the relationship depicted in the graph focused on the upper left hand corner of the graph because the errors of greatest concern are those where deleterious responses (above the lower horizontal line) are observed, yet the site falls below the lower nutrient threshold (left of the left vertical line). As can be seen in these graphs, the number of potential false negative assessment errors follows: GPP vs. TP = 2-3; ER vs. TP = 1; GPP vs. TN = 1; ER vs. TN = 1. UDWQ does not consider these errors to be evidence that the lower TN and TP criteria are not protective. If GPP or ER were always reliable measures of nutrient impacts, UDWQ would rely on these measures exclusively. The NNC do not assume that all moderately enriched streams will exhibit adverse effects for all responses. Instead, several responses are independently evaluated to identify adverse effects. The data support that the NNC are protective.

Of course, false positive impairment determinations—erroneous impairment conclusions—are also possible. In the context of use protection such errors are less problematic, but they can divert limited restoration resources. In the context of the graph provided by the commenter, these errors may occur for sites that fall to the right of the right vertical line, yet below the lower horizontal line. Sites in this region would be considered to be impaired wholly on the basis of nutrient enrichment, without evidence of adverse effects. However, each pane in the figure depicts a single response and one cannot conclude that this actually a false positive impairment determination until a similar conclusion is made for all NNC response variables. Nevertheless, the data presented in the graphs also suggest that the potential for false positive impairment determinations is likely to be relatively low.

Hypothetically assuming that these sites represent potential false positive decisions, the figure provided by the commenter demonstrates that both false positive and false negative error rates are low. This is one of the primary benefits of combining nutrient concentrations with ecological responses when setting NNC.

5B

In addition, a review of the GPP and ER results observed at Utah's headwater streams (blue symbols) shows that GPP and ER tend to be biased lower at headwater sites, at a given nutrient concentration, than the 2010 study sites. The TSD does not explain how the GPP and ER *threshold values developed using the 2010 sites (and not the 2015 headwater sites where these thresholds are applied) are protective of headwater streams aquatic life. This demonstration is key in supporting the proposed thresholds.*

UDWQ does not interpret these data to indicate a low bias. The responses were generally lower in the headwater streams than the 2010 statewide streams for similar TP and TN concentrations. As discussed throughout these responses and in the TSD, several known covariates that make streams naturally protective from the effects of nutrient enrichment are more prevalent in headwater streams. Lower responses relative to the TP and TN concentrations represent lower sensitivity to nutrients because of other covariates and are not indicators of bias. Many of the statewide streams from the 2010 sampling lack these protective covariates, and in these cases, the metabolic responses were expected to be higher. A failure to demonstrate a lack of adverse effects at streams with naturally protective physical attributes is not intrinsically problematic. In fact, such inconsistent S-R relationships formed the basis for using a combination of nutrient concentrations and ecological responses to assess moderately enriched streams.

5C

To further review the 2015 dataset, the EPA categorized each site based on the nutrient enrichment thresholds identified in Table 12.1 (page 139, TSD). Sites with TN and TP concentrations below the lower thresholds (<0.4 mg/L TN: < 0.035 mg/L TP) were labeled "low"; sites with either TN or TP concentration in the mid-range (TN= 0.40-0.80mg/L and TP=0.035- 0.080 mg/L) were labeled "moderate", and sites with TN or TP concentrations above the upper threshold (>0.80 mg/L TN; >0.080 mg/L TP) were labeled "high". Table 1 below presents the average and maximum values for TN, TP, GPP and ER for each "nutrient enrichment category." These data show that GPP does not clearly increase with increasing nutrient enrichment categories and ER decreases with increasing nutrient enrichment categories (which is the inverse of what we would expect if this indicator was responsive to nutrients). The lack of a predictable response in GPP and ER is problematic given the critical role the ecological response indicators play in the middle range of the TN and TP values for this combined criterion.

Table 1. Summary Statistics for 3 Nutrient Enrichment Categories in Headwaters Streams

Nutrient Enrichment Categories	Sample Size	Average TP	Average TN	Average GPP	Max GPP	Average ER	Max ER
low	20	0.02	0.24	0.99	3	2.69	9.83
moderate	20	0.04	0.35	0.52	1.51	1.32	4.91
high	9	0.13	0.53	1.24	3.37	0.73	1.53

The range of nutrient concentrations found among sites from all Utah streams (not just headwaters) demonstrated a relationship among ambient nutrient concentrations and metabolism responses (see Figure 5.2 above). The fact that the relationship weakens among streams with relatively lower nutrients is not surprising because there are fewer sites with the potential for nutrients to exert a measurable influence on metabolism and a relatively higher number of streams with attributes that are naturally protective from the adverse effects of nutrient enrichment. Second, particularly with respect to ER, the maximum values bias group averages. UDWQ discussed the site with the highest ER in the TSD and noted that this site also had the highest filamentous algae cover among all of the sites evaluated, so biological nutrient uptake is one logical explanation for the observed low nutrient concentrations at this site. Third, as the commenter previously noted, many of these streams have physical attributes that make them less susceptible to high GPP and ER rates, which when combined with more susceptible streams would weaken the relationship among metabolic rates and nutrient enrichment. The combination of nutrients and responses is intended to focus on the protection of aquatic life uses in streams at greatest threat to the adverse effects of nutrient enrichment.

5D

As the means to infer a linkage between these response indicators and aquatic life use protection, UDWQ examined the relationship between ER/GPP and DO. Specifically, the TSD documents the relationship between metabolic condition classes and the number of DO criterion exceedances (excursions below the 30-day average) as "independently derived indicators of potential threats to stream biota." The 2010 analysis showed an increase in DO excursions when ER > 5 g $O_2/m^2/d$. UDWQ performed this analysis using the 2010 statewide streams dataset (Figure 12.3, page 148, TSD) but not for the 2015 headwaters dataset.

The EPA ran the same analysis using the available data from headwater streams (n=30). The EPA's analysis showed no significant relationship between ER and the excursion of the 30-day average DO criterion -only two sampling locations had any probability of exceeding it (probability of 0.27 and 0.19). All other sampling locations showed zero probability of exceeding the 30-day average DO criterion of 6.5 mg/L. We also performed this analysis using the daily minimum DO value of 4 mg/L DO and only one site exceeded that value.

Based on the information in the TSD, UDWQ has not clearly demonstrated that GPP and ER are linked to protection of aquatic life uses in headwater streams as no linkage between exceedance of the DO criterion and GPP or ER (UDWQ's chosen metrics) was established.

The differences between the 2010 and 2015 investigations highlight the importance of using a sufficient range of ambient nutrient concentration for purposes of calculating ecologically meaningful response thresholds. ER is a mechanistically sound measure of conditions that lead to DO depressions and is more sensitive than DO. UDWQ interprets the lack of discernible relationship between ER and DO in headwaters stream as confirming the known insensitivity of DO for identifying adverse effects in moderately enriched streams. The 2010 data defined the relationship between ER and DO which was used to establish a protective threshold. The threshold is protective because it was based on streams that lack many of the protective covariates of the headwater streams.

EPA Recommendations and UDWQ Responses

The EPA recommends:

- *Providing documentation in the TSD that the proposed GPP and ER thresholds will ensure protection of aquatic life uses in headwater streams, particularly in light of the 2015 headwaters data. Specifically, the EPA recommends:*
    - *Conducting additional stressor-response analyses using only the 2015 headwater data and comparing the thresholds obtained with that analysis to the proposed thresholds for GPP and ER. Providing an explanation as to whether UDWQ evaluated the ER/DO relationship using the 2015 headwaters data and describe the results of that analysis.*

As discussed in the TSD (Chapters 1-7), the initial S-R investigation was intended to evaluate whether several proposed indicators of nutrient enrichment not previously collected were viable response parameters, and if so, whether thresholds derived from these parameters and existing data could be used to generate NNC for Utah streams. Given these study objectives, UDWQ collected data that encompassed the breath of ambient nutrient concentrations observed among Utah streams. The underlying assumption was that ecologically meaningful thresholds cannot be derived using data that were biased with respect to the known distribution of ambient nutrient concentrations. Ultimately, the results of this investigation were considered collectively to identify a combination of nutrient concentrations and responses that could be used to protect headwater streams from the effects of nutrient enrichment.

After proposing the NNC that combined nutrient and responses, the next step was an evaluation of whether or not the combination of nutrients and responses in the proposed criteria was adequately protective of headwater streams. As noted in responses to this section and Section 4, this confirmation exercise was conducted in 2015. Sites for this investigation were selected from those headwater streams that historical data suggested were moderately or highly enriched. Nutrients and responses were collected and UDWQ concluded that the combination of nutrients and responses in the NNC successfully identified those headwater streams where aquatic life uses were at greatest threat to nutrient enrichment (i.e., they had lower slopes or less channel shading than similarly enriched streams).

An alternative study design would have been to collect ambient nutrient concentrations and responses from as many headwater reference sites as possible to obtain a more robust estimate of background conditions. These data could then be used to better define the upper limits of naturally-occurring ecological responses. However, these data could not be used to determine whether departures from reference conditions constituted an adverse effect on aquatic life uses. As a result, this line of inquiry would not result in appreciably different thresholds or responses than those already proposed. To date, no stakeholder has presented defensible alternatives to the proposed NNC, despite all underlying data being made available. If alternative thresholds or responses were presented, UDWQ would be willing to consider alterations to the proposed NNC.

6. Other Potential Indicators

6A

UDWQ considered other response indicators in establishing the combined criterion for headwater streams, but the rationale for why they were not selected as response indicators is not clearly documented in the TSD.

Numerous candidate ecological responses were evaluated (TSD, Chapters 3-7), including: nutrient limitation, nutrient saturation, stream metabolism (GPP, ER), autochthonous organic matter standing stocks, diatom taxa composition, macroinvertebrate taxa composition, and macroinvertebrate biological assessment indices. In fact, the collective responses include virtually all of the responses recommended by an EPA expert panel that was convened to recommend the best ecological responses for purposes of assessing nutrient enrichment (EPA 2000). As described in the TSD and throughout the proposal, all responses were used as lines of evidence to for the primary nutrient thresholds to ensure that consideration was given for as many nutrient-related adverse effects as possible. As discussed in the TSD (pp. 145-151), this approach also allowed UDWQ to ensure that the proposed criteria included those responses that could be most easily incorporated into routine monitoring and assessment activities conducted by the agency and cooperators.

As discussed in the TSD (pp. 145-151), the proposed NNC ultimately included several responses that were the most direct measures of the two principal adverse effects of nutrient enrichment: increases in autotrophic and heterotrophic production. Other responses were not included in the proposed NNC because their routine collection was impractical (e.g., organic matter standing stocks) or because they are already routinely and independently evaluated (e.g., O/E, DO, pH). Metabolism, filamentous algae cover, and benthic algal biomass (for recreational uses) are all new responses incorporated into the proposed headwater NNC, and each has strengths and weaknesses. Beyond those already evaluated, the only additional response the UDWQ could potentially incorporate into the criteria would be diatom-based biological assessment metrics. UDWQ has not currently developed these assessment tools, so these would need to be incorporated once these tools have been developed and vetted with stakeholders.

6B

The TSD refers to Utah's analysis of potential biological indicators (i.e., diatoms, macroinvertebrates) to increasing nutrient concentrations which showed a loss of sensitive taxa at nutrient concentrations that were lower than the proposed lower nutrient thresholds. For example, the state's analysis shows loss of sensitive macroinvertebrate and diatom taxa at *approximately 0.011 mg/L TP and 0.016 mg/L TP which are much lower concentrations than the proposed TP threshold of 0.035. Similarly, the TSD states: "The overall threshold that captures the TP associated with the most appreciable changes in both tolerant and abundant diatom taxa was at 0.022 mg/L." (Page 89, TSD).*

This information suggests substantial changes to community structure, which is indicative of impacts to the aquatic life designated use, are occurring at lower levels than the proposed criterion, at least for TP. The TSD notes that:

"Allowing nutrient concentrations to achieve levels that result in assemblages dominated by tolerant taxa are likely underprotective." (Page 101, TSD)

Results presented in Chapter 7 (TSD) suggest that changes from sensitive to tolerant communities are occurring at nutrient concentrations below the proposed lower nutrient thresholds. This information suggests that the proposed combined nutrient criterion would likely lead to shifts to more tolerant, and less desirable, taxa – conditions that are not protective of aquatic life and that contradict the stated objectives.

The loss of highly sensitive taxa does not necessarily constitute an impairment to aquatic life uses. If numeric criteria are not violated, the point where departure from reference condition constitutes an impairment is a policy decision. It is widely acknowledged that an expectation of pristine environmental conditions is an unrealistic environmental objective. An acknowledgement of the importance of practical constraints in setting environmental objectives is intrinsic to all biological assessment programs.

TITAN identifies three thresholds: sensitive taxa, tolerant taxa, and taking both sensitive and tolerant responses into consideration (All). These thresholds provide additional lines of evidence with respect to TP concentrations that may be protective of aquatic life uses; however, they are one line of evidence among many and should not be interpreted in isolation of others. The comment is correct that the average threshold for the two most sensitive diatom responses is below the TP threshold proposed in the NNC. However, this is only true for the most sensitive of diatom taxa, which collectively revealed a threshold of 0.016 mg/L (95% confidence intervals = 0.010-0.022). Taking uncertainty into account, the proposed phosphorus threshold falls within the overall TITAN predictions: 0.022 mg/L (95% confident intervals = 0.010-0.047). Among all of the lines of evidence that UDWQ evaluated, these are the only thresholds that fall below lower nutrient thresholds. The tolerant taxa TITAN response threshold of .042 mg/L (95^{th} confidence intervals 0.027 – 0.051) was what UDWQ was referring to when making the statement about being under-protective (as quoted by the commenter), and this threshold is above the lower TP threshold proposed in the NNC.

The diatom TITAN analyses do suggest some statistically significant changes to this assemblage may occur at TP concentrations slightly below the proposed NNC thresholds, although ecologically significant effects are not predicted. TITAN thresholds only capture changes to those taxa that demonstrate statistically significant predictable changes along a stressor gradient. Many taxa did not change predictably with increasing TP, and presumably remain unchanged, so "substantial changes" at these concentrations are undetectable. The first significant change point for all other responses, including the macroinvertebrate O/E index that UDWQ uses for conducting biological assessments, falls above the proposed lower TP threshold. This suggests that the potential minor changes in diatom composition at the lower two diatom thresholds are not sufficient to alter the higher trophic levels or nutrient-related ecosystem functions. As noted in the TSD, the algae composition most likely to degrade uses is a regime shift from streams where benthic production is predominantly based on diatoms to one dominated by filamentous algae (See also responses to Section 6 comments) because these regime shifts often lead to habitat degradation and appreciable alterations of stream food webs. These types of changes are closely aligned with the protections afforded by Utah's aquatic life use descriptions, which specify that cold or warm water fish be protected, "including necessary aquatic organisms in their food chain" (UAC R317-2-6.3). Importantly, the lower diatom threshold for TP is below the 75th percentile of reference site TP concentrations which suggests a marked increase in false positive impairment determinations.

UDWQ has substantial diatom data and ultimately intends to develop biological assessment methods using this assemblage. Development of these analytical approaches will require sufficient resources. Once draft methods are available, it would be necessary to solicit stakeholder input before they are used to inform regulatory decisions. Once fully vetted, diatom responses could conceivably be included as additional responses indicators in a revised NNC. However, it is more likely that UDWQ would apply the methods independently, because (1) nutrients are not the only pollutants known to degrade diatom assemblages and (2) independent assessments would ultimately be more protective of designated uses.

EPA Recommendations and UDWQ Responses

The EPA recommends:
Exploring the use of diatoms or macroinvertebrates as response indicators for use with the combined criterion. Alternatively, the proposal should include a rationale for not selecting diatoms or macroinvertebrates as potential response indicators for the combined nutrient criterion.

As indicated on p. 88 in Chapter 7 of the TSD "Diatom compositional changes could only be evaluated for TP. TITAN revealed thresholds for eight diatom taxa that significantly decreased in abundance and occurrence in response to increasing TP concentrations (Figure 7.1)." While TITAN is useful for identifying response thresholds, the approach cannot be used to establish response thresholds that can be used for conducting assessments on streams not included in the analysis (see also response to comments in this Section). Diatoms were not selected because, based on the available data, thresholds could not be determined for TN, and because UDWQ has not currently developed diatom biological assessment tools.

Macroinvertebrate assessment scores could have been added as an additional response in the NNC, but DWQ opted to continue to keep these assessments independent due to the integrative nature of biological assessments. As discussed on p.

98 in Chapter 7 of the TSD, there was a significant relationship between nutrients and O/E scores: "Macroinvertebrate O/E scores decreased with increasing nutrients. A significant, albeit weak, linear relationship was observed among macroinvertebrate O/E scores and TN (n = 68, r^2 = 0.302, p < 0.001) and TP (n = 243, r^2 = 0.294, p < 0.001). The weakness in the relationship may be evidence of other stressors and natural gradients that are expected in structural responses because the effects of nutrients are indirect. For this reason, DWQ has incorporated functional responses with more direct linkages to nutrients into stressor-response (S-R) models directed toward NCC development." Macroinvertebrate biological assessments will continue to be conducted and the response indicators in the NNC will be integrated into biological assessment data collection activities. Keeping these assessments independent will allow UDWQ to better assess the relative influence of all stressors, including TN and TP, on the structure of stream biota.

7. Development and selection of the filamentous algal threshold

7A

The proposal and TSD currently lack a clear scientific rationale for the proposed 33% filamentous algae cover threshold proposed to protect aquatic life uses. The EPA's regulation at 40 C.F.R. § 131.11(a) requires that "criteria must be based on sound scientific rationale and must contain sufficient parameters to protect the designated use." UDWQ did not empirically derive the proposed filamentous algae threshold. Instead, the primary basis for the 33% threshold appears to be two studies available in the scientific literature (Biggs 2000, and Welch et al. 1988). Further examination of these papers reveals that the 30% threshold identified in Biggs is based on impacts to aesthetics and trout habitat in New Zealand.

The TSD does not include a description of which aspects of Biggs' conclusions are most relevant to Utah's headwater streams. For protection of benthic biodiversity, which should be protected as a component of the aquatic life designated use, it is important to note that Biggs establishes a mean monthly chlorophyll-a of 15 mg/m² as the threshold necessary -- not the 30% threshold cited in the proposal. In addition, Welch et al. 1988 relates to an aesthetics threshold for percent filamentous algae cover. Welch does not address aquatic life use protection. The proposal does not explain how a threshold based on aesthetics is protective of aquatic life uses.

It is true that one focus of these review papers identifies a decline in aesthetics and nuisance conditions degrading recreational uses. One possible reason is that recreational uses are likely more sensitive to filamentous algae growth than aquatic life uses. However, Biggs (2000) also discusses detrimental effects on aquatic life resulting from extensive filamentous algae throughout his review—loosely captured in summary tables as "aesthetics and trout habitat"—including: impairment of fish spawning, habitat degradation, reduction in fish feeding activity, reductions in fish populations, increases to pH and related increases in ammonia toxicity, and the potential for hypoxia. UDWQ is concerned that prolonged enrichment, particularly when combined with stabilization of flow via water diversions or impoundments, can cause streams to shift from diatom-based food webs to a state where the majority (i.e., >50%) of the stream bed is covered in filamentous algae. Such shifts have been observed in Utah streams and these prolonged periods of extensive filamentous algae blooms are considered to be an important indicator of eutrophication with a multitude of potential adverse effects to stream biota. UDWQ set an indicator that specifies that filamentous algae shall not exceed ⅓ cover to be protective against dominance of this growth form (see also response to section 3B). This section of the response to comments summarizes some of the important linkages to aquatic life uses discussed by Biggs (2000) and UDWQ has also included additional information on the linkage in the TSD.

Extensive and long-lived filamentous algae blooms can degrade stream habitat in several ways. First, the structure of these algae mats are known to trap suspended sediment, which can fill interstitial spaces in cobble-bedded streams. This diminishes the quantity and quality of macroinvertebrate habitat because they frequently reside in these interstitial spaces. Extensive algal mats also slow the flow of water, which can increase stream temperature in areas exposed to sunlight. Increases in water temperature can have adverse metabolic effects on macroinvertebrates because these organisms are ectotherms, with body temperatures dictated by their surrounding environments. Higher stream temperatures also decrease DO saturation, which can be detrimental to stream biota.

Another adverse effect of a regime shift to a condition where filamentous algae blooms are extensive (i.e., >50% cover) is the alteration of stream food webs. For most stream macroinvertebrates, diatoms are a preferable food resource because they are higher in lipids and more easily digestible than filamentous algae. Some algal piercing taxa are specialized in feeding on filamentous algae and they will replace less specialized taxa when algal mats are present. However, these taxa are also less likely to actively drift, which can be detrimental to the feeding efficiency of trout and other fish. Decreases of trout abundance in streams dominated by filamentous algae have been documented and this may be due, in part, to shifts in the taxonomic composition of lower trophic levels. In mesotrophic streams, the loss of actively drifting stream insects is less severe because algal mats are more short-lived, allowing time for recolonization of these insects. Another impact of excessive filamentous algae growth is a reduction in the grazing efficiency of macroinvertebrates. Top down control of algal biomass through grazing has been well-documented in streams where stream autotrophs consist primarily of diatoms (the natural condition of Utah headwater stream), but these biotic controls on excessive benthic production are less effective once filamentous algae are the dominant forms of these ecosystems.

Extensive mats of filamentous algae also exacerbate other water quality parameters that have been demonstrated to be deleterious to aquatic life uses. Higher rates of primary production are associated with increases in pH, with the potential to negatively affect the condition of fish and indirectly affect all aquatic biota by increasing the toxicity of ammonia that may be present at higher concentrations in nutrient enriched streams. The resulting higher biomass of autotrophs can cause appreciable increases in the diel flux of DO, which has been attributed to biological degradation. The accrual of autochthonous carbon in streams with extensive filamentous algae mats can be considerable, especially when nutrient concentrations are high because episodic sloughing decreases and filamentous algae mats become longer and thicker due to increases in nutrient diffusion into the mats. In Utah, decreases in stream temperatures will ultimately result in algal senescence, which allows the carbon tied up in these mats to be more available to stream heterotrophs, increasing ER, which has the potential to cause hypoxic conditions (Suplee et al. 2019). Another advantage of measuring filamentous algae cover during the growing season as a nutrient response is the ability to identify the potential for late-season deleterious hypoxia.

The natural state of headwater streams in Utah is a state where stream autotrophs primarily consist of diatoms. UDWQ included filamentous algae as a response variable based on the knowledge that regime shift to autotrophs dominated structurally in filamentous form is undesirable, from the perspective of both recreational and aquatic life uses. Existing scientific evidence establishes clear linkages between nutrient enrichment and adverse effects on <u>both</u> of these uses.

As this review illustrates, enough is understood about the deleterious impacts of extensive filamentous algae blooms to reasonably conclude that dominance of this growth form in headwater stream constitutes a threat to aquatic life uses. A more precise estimate of exactly how high seasonal maximum algae growth can go before these adverse effects occur is limited by the available data. As Biggs (2000) states, "*effects on water quality and ecosystem degradation are only moderately well quantified, and a number of cause-effect assumptions in this linkage need further testing.*" It would be useful for EPA to take a leadership role in helping to fill these data gaps because filamentous algae cover is an important nutrient response that could be easily and inexpensively be integrated into state monitoring programs.

7B

Beyond the literature citations, the proposal provides the following justification for selecting the 33% threshold:

"DWQ recommends a criterion of maximum filamentous algae cover of 1/3 of the stream bed. While this number is at the upper end of concentrations that others have suggested as protective of stream aquatic life uses, <u>DWQ has established this threshold as protective of stream conditions because it represents the maximum filamentous algae concentration that is observed on any single collection event</u>." (Proposal, page 27) [emphasis added]

Given the stated paucity of data collected on Utah streams, there is limited information on which to base this statement and demonstrate that the proposed threshold is protective of aquatic life uses in headwater streams.

The derivation of the filamentous algae NNC threshold was semi-quantitative. UDWQ has provided evidence to support that nutrients must be controlled to prevent filamentous algae from becoming the dominant form of benthic autotrophs, which is defined as >50%. As discussed elsewhere in responses to this section, this is because extensive algal blooms are more likely to be associated with nutrient enrichment and also because they represent a greater threat to aquatic life uses. Given this determination, any value less than 50% is considered to be protective. Filamentous algae blooms can occur in relatively unenriched streams with stable flows, but they tend to have a more limited spatial and temporal extent. To avoid making too many false impairment determinations due to natural filamentous algae accumulations, UDWQ used 20% to demarcate an upper limit of naturally-occurring conditions. UDWQ met with a technical review team and the consensus was that a NNC response to maintain filamentous algae cover lower than ⅓ of the stream, a value approximately midway between 20-50%, was the most appropriate way to balance false positive and false negative decision errors. The supporting documents have been updated with additional clarification

7C

The EPA's guidance document entitled "Guiding Principles on an Optional Approach for Developing and Implementing a Numeric Nutrient Criterion that Integrates Causal and Response Parameters" recommends that "[combined] criterion should demonstrate the sensitivity of the response indicator(s) to increased nutrient concentrations and quantify how these nutrient-response linkages will achieve the goal of protecting and maintaining aquatic communities." Chapter 13 of the TSD, page 160, explicitly states that "there was no statistically significant linear relationship between TN or TP and filamentous algae cover." The proposal also cites to the influence of confounding factors on filamentous algae growth: "Others have noted that whether or not filamentous algae cover reaches levels of potential concern is also dependent on other stream characteristics such as canopy cover, stream temperature, stream size, and hydrology (Busse et al. 2006, Dodds and Oakes 2004, Riseng et al. 2004)." (Proposal, page 27). Based on the proposal, it is unclear how UDWQ controlled for confounding factors in the study design used for the independent evaluation of the proposed numeric nutrient criterion (Chapter 13) or in any subsequent analysis.

The quantity of filamentous algae is mechanistically related to nutrient concentrations, although this relationship was not apparent in the available data. The proliferation of filamentous algae is among the oldest biological indicator of nutrient enrichment in streams. As streams become enriched, it is increasingly likely that proliferations of filamentous algae will occur. UDWQ interprets a dominance (>50%) of filamentous algae to be a strong indicator of nutrient impairment with a low potential for false positive impairment determinations. The converse interpretation, concluding that a stream without filamentous algae is in an oligotrophic state, is not an appropriate interpretation of the data. As highlighted in the TSD and other responses to comments, local stream characteristics exert controls on algal accrual. Flooding frequency is particularly important because scouring can reset algal accrual. Hence, it is difficult to reliably link low levels of filamentous algae to oligotrophic stream conditions. To reduce the probability of making these false negative impairment determinations based on filamentous algae, the thresholds in the NNC are expressed as "not to be exceeded" values, with a minimum recommended monitoring frequency of once per month. The implementation of filamentous algae at sites with moderate levels of nutrient enrichment is not intended to address the situations where natural factors prevent proliferations of extensive algae growth. However, when present, extensive filamentous algae cover is, in and of itself, a reflection of degraded aquatic life uses.

7D

The EPA also reviewed regionally-relevant studies related to Utah's proposed filamentous algae threshold. For example, the Montana Department of Environmental Quality (DEQ) conducted a dose-response nutrient study on a perennial prairie reference stream. Recognizing that plains streams, even reference sites, generally tend to have a higher degree of human disturbance than montane reference sites (i.e., "least disturbed" compared to "minimally disturbed" reference sites), the EPA examined the Montana study because it offered insights into visually assessed filamentous algae values observed across a nutrient gradient. The EPA's hypothesis was that the visual estimate of filamentous algae observed in prairie reference streams would be greater than estimates from headwater streams because prairie streams are likely to be less shaded, have lower gradients, more impacts, etc. Therefore, the Montana study represented the "upper range" of conditions for filamentous algae that could be applicable to headwater conditions.

UDWQ again reviewed the findings of the Montana study in response to this comment. Based on the information provided in this study, UDWQ does not support that the prairie streams are closely representative of the "upper range" for filamentous algae in Utah headwater streams. Prairie streams have some physical attributes that potentially make them more susceptible to nutrients. However, these streams also typically have smaller substrate size, particularly in slower runs and pools, which tends to favor macrophytes over accumulation of filamentous algae. This is important because the investigators note that most of the filamentous algae growth occurred in riffles, which tend to have larger, more stable substrate. The vast majority of Utah's headwater streams have larger, more stable substrates, which make accumulations of extensive algal mats more likely. Moreover, on average, riffles only comprised 15% of the stream reach, further limiting the spatial extent of filamentous algae mats observed in the Montana investigation. Again, Utah's headwater streams have a higher proportion of stream reaches in riffles and runs, which makes direct comparison with the Montana investigation problematic. As reported in the TSD, UDWQ collected filamentous algae data at 49 headwater streams located throughout Utah. Filamentous algae cover at these streams peaked at 95% cover, with 16% of streams exceeding the 1/3 threshold proposed by the headwater NNC. Clearly the filamentous algae cover data reported in the Montana study is not reflective of the "upper range" applicable to Utah's headwater streams, because empirical data collected from Utah headwater streams demonstrate that more extensive filamentous algae mats occur in the streams the NNC are intended to protect.

Another interesting observation from the Montana investigation is that it demonstrated many of the same deleterious responses discussed elsewhere in these responses to comments. This investigation clearly demonstrated that filamentous algae cover increases in response to nutrient enrichment. The stream that received the higher nutrient dosing had much higher daily maximum DO concentrations in comparison to the control, and modestly lower daily minimum concentrations, which suggests increases to both GPP and ER rates, which the researchers link to degradation of aquatic life. The high dose stream also exhibited decline in DO below water quality standards, which the authors of the investigation attributed to the respiration of the additional carbon during the senescence of filamentous algae. In their words (Suplee et al. 2019, emphasis added), "*A consistent finding across the whole-stream enrichment literature is that increased nutrients increase algal primary productivity and standing crop of a stream and, accordingly, we monitored the floral community. Enrichment effects on the benthic algae of this prairie stream were consistent with this nearly universal finding, and the degree of changes was largely in alignment with our dosing levels.*"

7E

Lastly, visual estimates for filamentous algae cover can be prone to bias. Therefore, having sampling methods that ensure a systematic approach to sampling is important. There is also no information describing how the proposal evaluated bias in the visual estimates.

Detailed standard operating procedures (SOPs) were written prior to data collection and all field crews were trained together to ensure that they had interpreted the methods similarly. To address the potential for sampling error, the SOP requires measurement to be taken from three transects running perpendicular to stream flow and located in riffles and runs—where filamentous algae cover is more likely to occur. Along each transect, 10 measurement of filamentous algae cover are made upstream and downstream in each of 4-6 habitat units. Each measurement of the presence or absence of filamentous algae is made using a line intersect method, which is a method commonly used to quantify plant cover in vegetative surveys. This protocol resulted in >240 measures of filamentous algae cover from each stream, which is a sample size that should be more than sufficient to minimize observation bias.

EPA Recommendations and UDWQ Responses

The EPA recommends:

- *Adding a chapter to the TSD that describes filamentous algae as a response indicator and addresses the issues described below;*

  UDWQ added some of the material from the responses to this Section to Chapter 12 of the TSD.

- *Discussing in the TSD how the scientific studies cited demonstrate that a threshold of 33% filamentous algae cover will protect aquatic life uses;*

  Additions to the TSD include a more thorough description of the underlying rationale for the selection of the 33% threshold (see also response to Section 7B).

- *Explaining in the TSD how UDWQ's study design controlled for confounding factors in the derivation of the filamentous algae threshold;*

The revisions to the TSD include a more thorough description of the underlying rationale for the selection of the 33% threshold (see also response to Section 7B).

- *Reconciling the results from Montana's dose-response study with the proposed 33% threshold and demonstrating in the TSD that the threshold adequately protects aquatic life uses;*

Please see response to Section 7D. UDWQ did not revise the SOP to incorporate the information because the study is not directly applicable and the resulting arguments only further support the information already presented, as modified, in the TSD.

- *Providing a copy of UDWQ's standard operating procedures (SOPs) for determining visual estimates of percent filamentous algae growth;*

The SOPs were developed prior to collection of the data, revised after being vetted in preliminary field testing and will be submitted in conjunction with the rulemaking packet provided to EPA for formal review and approval in accordance with section 303(c) of the CWA.

- *Describing how the state evaluated potential bias in the visual estimates;*

Bias is minimized through the use of collection techniques that are easily repeatable. UDWQ used a derivation of line-intersect, transect measurements that are widely used to quantify plant cover in vegetative surveys. To ensure the methods are repeatable, detailed standard operating procedures (SOPs) are provided and all field crews were trained together to ensure that they had interpreted the methods similarly. UDWQ will also work with cooperators interested in collecting these types of data to minimize the potential for bias. Another important factor that limits the bias of any measurement is sample size. As stated in the TSD, filamentous algae cover was calculated from >240 field measurements, which were collected on each collection event.

- *Characterizing the percent filamentous algae cover at reference streams in the TSD; and*

Due to the confirmational nature of the 2015 investigation, which focused on data collection in moderate to highly enriched headwater streams, UDWQ does not have the data necessary to conduct the requested analysis. However, UDWQ agrees that a thorough investigation of background filamentous algae cover would be useful to the scientific community and will endeavor to help fill this important gap as additional data become available through implementation of the NNC.

- *Exploring the relationships between percent cover and invertebrate data.*

There are currently insufficient co-located samples to evaluate these relationships. However, UDWQ is integrating collection of filamentous algae cover into routine biological assessment data collection activities, so these analyses can be conducted in the future.

8. Derivation of lower and upper nutrient thresholds

The proposal applies a combined nutrient criterion within a broad range of TN and TP values - ranging from concentrations representative of reference conditions to upper TN and TP values at which impairments to aquatic life are expected. Given the EPA's comments on the sensitivity of the response indicators and protectiveness of the proposed thresholds, UDWQ's proposed approach could allow nutrient concentrations to increase until they reach the upper nutrient thresholds and impact or impair aquatic life.

As discussed in these responses, UDWQ supports that the combined criteria used to evaluate waters moderately enriched by nutrients are sufficiently protective because of the multiple lines of evidence. The potential for false negative impairment determinations is further limited in occurrence and extent by having the upper thresholds. However, the primary purpose of the upper thresholds is to conserve monitoring resources by rapidly and efficiently identifying impaired sites and to protect downstream uses.

EPA Recommendations and UDWQ Responses

The EPA recommends:

- *Explaining how a broad range of nutrient concentrations will ensure protection of aquatic life uses.*

Chapter 12 of the TSD has an extensive discussion of the rationale behind the NCC and evidence that the combination of nutrients and responses is protective of aquatic life uses. The response parameters included in the TSD are only applicable to moderately enriched streams. The NNC specify that a violation of any response in moderately enriched streams is sufficient to consider a stream to be impaired. If adverse effects are not observed at moderately enriched streams, it likely means that the stream has physical attributes that reduce its sensitivity to nutrient enrichment. Under such circumstances, there is little reason to believe that moderate enrichment will "impact or impair aquatic life." The commenter does not present any data that demonstrates that the combination of nutrients and responses in the NNC miss adverse effects of nutrient enrichment on aquatic life uses.

- *Providing documentation on the reference site screening process.*

The reference site selection process was developed prior to collection of the data and will be submitted in conjunction with the rulemaking packet provided to EPA for formal review and approval in accordance with section 303(c) of the CWA.

- *Clarifying in the TSD and proposal, if needed, the number of reference samples and whether the 90th percentile represents a mean concentration.*

The TSD (Chapter 12) already summarizes the number of samples used for calculating growing season averages (see *Data Compilation* section in *Methods*). In addition, data used to conduct the analysis of reference sites were made available to stakeholders, including EPA. The NNC specify that nutrient concentrations are to be based on the average concentration of TN and TP over the growing season. As a result, UDWQ treated the reference site data similarly. For each reference site, the average of all available samples, collected during the growing season, over the nine-year period of record, were calculated (see Table 11.1). The averages among all reference sites were then used to determine percentiles (see Table 11.2) . The TSD and Proposal have been updated to more clearly state the percentiles were derived from growing season averages.

9. Full Support Assessment Determinations Based on a Single Response Indicator

As described on page 2 of this letter, UDWQ will in some instances be relying on data from a single response indicator in use attainment decisions. Page 42 (Proposal) states that sites will be considered "to be meeting their aquatic life uses provided that at least one response variable has been measured and no response that has been measured exceeds the established thresholds…In circumstances where a response is required to make an assessment decision, it is not necessary to have data on all response[s] specified in the NNC."

For states using a combined criterion, the EPA recommends relying on data for more than one response indicator to make the decision that a water is fully supporting its aquatic life uses. This approach is especially critical in situations where the response indicators, and their associated thresholds, may not be very sensitive to increases in nutrient concentrations.

Based on the EPA's technical review of the proposed suite of response indicators, none of Utah's proposed response indicators exhibit a strong relationship to nutrients. In addition, because of an inability to calculate GPP and ER at many headwater sites (see page 5 of this letter), the state may have to rely on results for percent filamentous algae cover to make attainment decisions based on a threshold that lacks a clear scientific rationale. Therefore, the EPA recommends modifying the proposal to require that data from all response indicators must be collected and meet the proposed thresholds before a site can be assessed as meeting its aquatic life uses.

While EPA guidance of combined criteria recommends the inclusion of an assessment matrix, this is not a requirement of the CWA and associated regulations. UDWQ typically provides specifics with respect to the interpretation of standards in the Integrated Report's Assessment Methods. In this case, the agency agreed that an exception was appropriate due to the nuances intrinsic to combined criteria. The assessment decision rules that are part of the NNC were policy decisions made in consultation with stakeholders.

The circumstances described by the commenter are those where either TN or TP indicate moderate enrichment and evidence exists that adverse effects do not occur in the stream. Importantly, in such circumstances there would also be no evidence that deleterious responses have occurred. UDWQ is comfortable making a full support determination in these circumstances unless other nutrient-related criteria (e.g., pH, DO) identify adverse effects to aquatic life uses. The decision to consider such sites to be fully supporting aquatic life uses was made to maximize agency resources. The alternative to a full-support listing decision would be assessing the site as having insufficient data (Assessment Category 3), which obligates the agency to collect follow-up data. Given that collection of response data is somewhat resource intensive, prioritization of data collection at these low-nutrient sites over those that are moderately enriched streams or without any nutrient data is not sound planning. More importantly, such diversions of limited resources would diminish the ability of UDWQ to ensure the statewide protection of all headwater streams. As demonstrated by the 2015 pilot investigation, the proposed NNC will allow UDWQ to identify headwater streams at greatest threat to nutrient enrichment.

EPA Recommendations and UDWQ Responses

Therefore, the EPA recommends modifying the proposal to require that data from all response indicators must be collected and meet the proposed thresholds before a site can be assessed as meeting its aquatic life uses.

The decision rules were policy decisions that were made in consultation with stakeholders. UDWQ intends to routinely collect data to support all of the ecosystem responses to improve the accuracy of the assessment decisions. However, each individual ecosystem response has been demonstrated to be sufficiently robust to support an assessment decision of full support unless other water quality indicators demonstrate adverse effects. It is also worth noting that the converse is also true; the NNC specify that an excursion of a single response threshold is sufficient to make a determination of impaired conditions. The assessment decision rules in the proposal have not been modified.

10. Criterion Operation and Implementation Comments

In addition, the EPA offers the following set of criterion operation and implementation-related comments for UDWQ's consideration:

- *We recommend developing SOPs for making assessment decisions within the combined criterion matrix to clearly document how assessment decisions will be made.*

UDWQ revised the TSD and Proposal to provide clarification with respect to how assessments will be conducted. UDWQ previously documented how the criteria will be implemented for assessments in Tables 4 and 8 and accompanying text in the proposal. In addition, Table 8 is explicitly incorporated into the rule. Additional ongoing clarifications can continue to be provided in the methods associated with the *Integrated Report*.

- *Because the combined criterion incorporates a range of nutrient thresholds, please describe in the proposal the process that the state would use to identify total maximum daily load (TMDL) targets when a waterbody is listed as impaired.*

UDWQ has not developed implementation methods for TMDLs because TMDL targets will be based on achieving the criteria. UDWQ considers both the lower and middle tiers equally to be supporting beneficial uses and acceptable as TMDL targets. TMDL targets will therefore depend on site-specific relationships between nutrient concentrations and ecological responses. Detailed TMDL guidance will be developed in consultation with stakeholders prior to the NNC being used to make impairment determinations in the *Integrated Report*.

- *In the TSD and proposal, please clarify the assessment decision that would occur if TP concentrations exceed the moderate or upper thresholds but TN concentrations do not.*

A minimum of 4 samples are required to calculate growing season TN and TP averages for assessment purposes. Broadly, the NNC intend final assessment decision would be made based on the extent to which nutrient enrichment threatens aquatic life, so DWQ would base its overall assessment on the worse-case scenario. Table 2.14.8 in R317-2 shows the criteria, and if all of the required conditions within a tier are met, the criteria are being met. Therefore, a site must meet both TN and TP thresholds to be considered meeting the tier-specific criteria. As specified in Table 2.14.8, if TP concentrations exceed the lower threshold but TN concentrations do not, ecological response data are required to evaluate the TP exceedance. If TP concentrations are above 0.08 mg/L, the site is considered to be not meeting the criteria, regardless of TN concentrations.

- *Please clarify in the TSD and proposal whether both TN and TP concentrations need to be below the lower thresholds for a site to be considered fully supporting its aquatic life uses. The current language in the proposal appears to offer contradictory perspectives. For example, the description in the Proposal (pages 24, 41) specifies that both TN and TP must be below the lower threshold for a site to be considered meeting its aquatic life uses. In contrast, Tables 4 and 8 (pages 33 and 42, Proposal) suggest "either/or" must be below the specified nutrient concentration before a full support determination can be made. The TSD (pages ES7, 137 and Table 12.1 on page 139) contain similar contradictory information.*

The requirements specified in rule are consistent and clear. As required by proposed Table 2.1.4.8 in R317-2, both TN and TP concentrations must be below the lower thresholds for a site to be considered fully supporting as indicated by the "and."

- *The EPA recommends UDWQ consider placing waters identified within the moderate range of TN and TP, with no available ecological response indicator data, in Category 5* *of the Integrated Report until response indicator data are available to confirm the aquatic life use is attained.*

As shown in Table 12.1 of the TSD, waters without any response indicator data will be assessed as having insufficient data: "Streams without response data will be listed as having insufficient data and prioritized for additional monitoring if either TN or TP falls within the specified range." UDWQ's approach is consistent with the EPA's 2013 Guiding Principles for biocriteria that state: "If a causal parameter is significantly exceeded but no response parameters are exceeded, then the state should pursue additional studies to determine whether site-specific criteria are appropriate."

- *Please describe in the proposal how the proposed combined criterion will ensure protection of downstream uses, especially in instances where nutrient concentrations are allowed to increase to/near the upper thresholds for TN and/or TP.*

Implementing these criteria will be a vast improvement to protect and restore downstream uses considering that mid-to lower elevation streams currently lack numeric nutrient criteria. In most cases, the most sensitive downstream uses are likely to be lakes or reservoirs. UDWQ currently assesses these waterbodies using several nutrient-related responses, including: trophic state index (TSI), DO, pH and the relative abundance of cyanobacteria. In addition, UDWQ conducts biological assessments in locations downstream of headwaters. These biological assessments have been documented in the TSD to be sensitive to nutrient enrichment. Any impairments that are identified would be addressed through TMDLs, which may require nutrient reductions that are more stringent than otherwise needed for the prevention of adverse effects in headwater streams.

APPENDIX B: ADDITIONAL MATERIALS IN SUPPORT OF MECHANISTIC MODELING

This appendix contains reports that were submitted in conjunction with the modeling component of the nutrient study. The studies primarily speak to the integration of process-based models in the creation of site-specific criteria, but some information is of broad interest.

A Data Collection and Calibration Strategy for QUAL2Kw

A.J. Hobson,[1] B.T. Neilson,[1] N. von Stackelberg,[2] M. Shupryt,[2] J. Ostermiller,[2] G. Pelletier,[3] S. C. Chapra[4]

1. Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University

2. Utah Department of Environmental Quality, Division of Water Quality

3. Washington State Department of Ecology

4. Civil and Environmental Engineering, Tufts University

Abstract

In-stream water-quality models provide guidance in watershed management decisions by linking pollutant loads to changes in water quality. These models are particularly useful for determining wasteload allocations, developing numeric nutrient criteria, and aiding in total maximum daily load (TMDL) analyses. Unfortunately, the routine data collected as part of the governmental monitoring efforts do not typically meet the data requirements for modeling. Consequently, this study presents a foundational data collection methodology suited to meet in-stream water-quality modeling requirements for a commonly used model (QUAL2Kw). To set some model parameters directly, methods are provided for estimating maximum sediment oxygen demand and appropriate reaeration formulas using observed oxygen time series. The quantity of many data types was minimized to reduce cost which resulted in challenges due to data limitations (e.g., designation of appropriate loading values from highly variable point source information). Similar to other modeling studies, parameter estimates were also not readily identifiable. However, even simple methods to reduce the number of parameters requiring calibration proved beneficial. Although most problems will require additional model calibration and data for model corroboration, this approach provides an initial framework that aids in the judicious use of resources to meet watershed management decision making needs within the context of wasteload allocation and/or numeric nutrient criteria development.

Introduction

In-stream water-quality models can be helpful in the watershed management decision process by understanding nutrient loading effects on changes in water quality (Boyacioglu and Alpaslan 2008; Kannel et al. 2011; National Research Council 2007; Orlob, 1992; von Stackelberg and Neilson 2014). Such models are used for a variety of applications including wasteload allocations (WLAs) (UDEQ 2012b), establishing regional or site-specific numeric nutrient criteria (NNC) (Flynn and Suplee 2011; US EPA 2000), and total maximum daily load (TMDL) assessments (Boyacioglu and Alpaslan 2008; National Research Council 2001). Many of these applications focus on critical low-flow periods (Bischoff et al. 2010; Gunderson and Klang 2004; Stahl and Smith 2002; UDEQ 2000; US EPA 2002a) that result in high primary productivity, low dissolved oxygen (DO) levels, and elevated stream temperatures (US EPA 1997). These conditions often exceed in-stream water-quality standards, approaching thresholds of many aquatic organisms (Hester and Doyle 2011), and are only expected to worsen in the future with global climate change (Whitehead et al. 2009). During critical periods in riverine systems, simplified one-dimensional, quasi-dynamic (constant flow with diel weather and water quality) models, such as QUAL2E (Brown and Barnwell 1987) and QUAL2K (Chapra et al. 2008), are typically employed to represent the fate and transport of solutes in the downstream direction (US EPA 1997). A modified version of these models maintained and distributed by Washington State Department of Ecology, QUAL2Kw (Pelletier et al. 2006), has been selected for water-quality impairment assessments conducted by many state and national agencies (Carroll et al. 2006; Kannel et al. 2011; Turner et al. 2009).

All models require data input for model setup, including physical characteristics (hydraulic information and channel segmentation), forcing (meteorological, boundary conditions, and point and distributed sources) and calibration data (in-stream observations) to adequately characterize effects on water quality from significant loading sources. The supporting data collection campaigns must capture stream variability (both temporal and spatial) and necessary data types while often adhering to stringent budget requirements (Neilson and Chapra 2003; US EPA 2002b). As the need for WLAs, TMDLs, and site-specific NNC causes a burden on internal resources for public and private agencies (Lettenmaier et al. 1991), a reliable and systematic method of collecting data to support in-stream modeling is needed. Sampling strategies must be established that limit the number of measurements and data types collected without having a consequential impact on model reliability (Dunnette 1980; Facchi et al. 2007; Henderson-Sellers and Henderson-Sellers 1996).

To address the need for protocols to optimize the allocation of limited resources, this paper presents a systematic and foundational data collection, model setup, and model calibration framework for applying QUAL2Kw to riverine systems impacted by water reclamation facilities (WRF). Assuming steady state flow conditions, the generalized approach provides guidance for basic data collection given the temporal and spatial variability of point source impacted reaches, as well as options for additional data collection to support model parameterization and calibration. To evaluate the effectiveness and limitations of the generalized and basic approach, a case study is presented of a model application to an effluent-dominated headwater stream in Utah.

Generalized Data Collection and Modeling Approach

The generalized approach was initially developed and applied to 6 study sites throughout central and northern Utah (SI Figure 1). Throughout this process, site specific hydrologic characteristics, sampling problems (e.g., missing samples), and varying influences of the WRFs provided insight regarding data requirements and provided the basis for the sampling plan presented here. This generalized data collection and modeling framework relies upon synoptic surveys of point source impacted reaches to support model setup and calibration. Synoptic surveys require concurrent sampling of the effluent as well as upstream and downstream locations. These surveys are especially important during the critical period, which is typically summer low-flow conditions for this type of system (e.g., Turner et al. 2009). However, additional synoptic surveys collected under various environmental conditions would serve for model confirmation after calibration, and also to define the temporal extent of the critical season to guide seasonal wastewater treatment requirements.

Sampling Locations

To model a study reach (Figure 1), data must be collected to capture the variability of the headwater (also called the upstream boundary condition, Station 1), point sources or abstractions (e.g., WRF, tributary inflows, irrigation diversions at Station 2, T1 and D1, respectively) and any diffuse sources or abstractions (e.g., groundwater). The type of information that should be gathered at each station varies (Figure 1, Table 1). In the context of effluent dominated systems, at an absolute minimum, supporting data need to be gathered at the headwater/upstream boundary condition (Station 1), point source before it enters the stream (Station 2), and the downstream end of the study reach for calibration (Station 3). However, placing additional stations at the initial mixing point of a WRF and at or beyond the point of maximum impact of the point source is desirable. If a significant tributary (e.g., greater than 10% of the study reach flow (Bartholow 1989)) enters the modeling reach, flow and quality data must also be collected at Station T1. Also, if a significant diversion is present, flow information is needed at Station D1 (the water quality of the diversion is the same as that in the modeled reach).

Once the stations are identified, the distance between Stations 2 and 3 can be determined using various methods and informed by site-specific criteria, but should located downstream of the mixing zone and be long enough to capture the processes of interest. In general, selection of the Station 3 location must balance the need to capture the maximum effect of the discharge while minimizing confounding factors of tributaries, diversions, and groundwater. Further, it may be appropriate to select the distances based on requirements to derive estimates of open-water metabolism and surface reaeration (k_a) which may require additional intermediate stations.

Data Types

Data are required for a number of water-quality constituents at each station. The requirements are dependent on whether it is the headwater station, a point or distributed inflow, or a diversion (Figure 1, Table 1). Data collected at each station should represent the study period of interest (e.g., summer low flow conditions). Some constituents can be sampled directly while others are estimated using relationships between measured constituents and model variables (SI Table 1). An estimate of bottom or benthic algae concentrations should also be measured, particularly in shallow streams or rivers. Additional data types that could be collected include a measure of sediment oxygen demand (SOD), total organic carbon, and dissolved organic carbon (DOC) (used to estimate CBOD and/or detritus).

Beyond water-quality data, site-specific information is necessary to characterize the stream and its surroundings. This includes geometric (bottom slope, channel cross-sections), meteorological (air temperature, dew point temperature, wind speed, and cloud cover or solar radiation), and hydraulic (travel time, stream and groundwater flow, velocity, substrate types, and percent suitable substrate) information. A summary is provided with a list of the data types to collect, some procedural information, locations where these data are required within or near the site, and the utility of the data in the context of the modeling effort (Table 2).

Sampling Frequency

When one considers the timing of sample collection in the context of point source impacted reaches, it can be difficult to resolve whether observed diel fluctuations are from point source variability or simply an artifact of sampling time and background diel fluctuation cycles (Nimick et al. 2011). When possible, site-specific information on spatial and temporal variations of specific water-quality constituents should be obtained prior to committing to a sampling plan (Ort et al. 2010). For example, pre-sampling reconnaissance can include deployment of continuous sondes to explore the variability and timing of diel minimum and maximum conditions at various locations. In general, those constituents that have the highest variability require the most frequent sampling interval. Some constituents that can be measured *in situ* using multi-parameter sondes (e.g., temperature, DO, conductivity) can be sampled most frequently during the study period. Although these sensors are typically limited more by cost than by temporal sampling frequencies, measurements made hourly for at least a 24-hour period, and preferably over 2–3 days or the duration of the synoptic sampling event, should adequately capture a typical diel signal and provide appropriate estimation of constituents for modeling needs (Gammons et al. 2011). Grab sample frequency of the remaining constituents requiring laboratory analyses are typically limited by personnel and cost. Therefore, entities commonly rely on intermittent sampling due to assumed low diel variability or on historic values for modeling applications (Bischoff et al. 2010; Carroll et al. 2006). Sampling frequency has been studied extensively (Facchi et al. 2007; Fogle et al. 2003; Hazelton 1998; Henjum et al. 2010; Ort et al. 2010; Zhang and Zhang 2012) and some guidance on choosing a sampling strategy is offered specifically for applications in general TMDL analyses and WLAs (US EPA 1986, 1995, 2002b). If possible, sampling should occur at least twice a day to target the times of expected diel minima and maxima of various constituents (e.g., at dawn and after solar noon or dusk) during the beginning and end of the study period at all key stations (Chapra 2003).

Model Setup

Once these water-quality and supplemental data are collected, model setup requires translation of stream observations to the input format requirements of the model. First, the study reach must be segmented and information regarding the channel geometry of each model reach must be determined from the observations. Next, all flow records from each station should be averaged to provide a single flow value for each point source/diffuse inflow, tributary, diversion, and the headwater. Then, measured water-quality data needs to be averaged to produce hourly estimates (headwater), or summary statistics (average, min, max, time of max concentration for assumed sine curve) for point source loads. In the case of limited data availability, as is common among chemistry and nutrient sampling, values can be averaged to provide daily mean concentrations and applied as a single value that does not have diel variation.

Another consideration with limited data availability occurs when chemistry and nutrient samples are analyzed and reported at very low concentrations that result in censored values or samples reported as below analytical or method detection limits (MDL). These cases require consideration of appropriate methods for estimating the true values, since either omitting a censored value, replacing it with zero, 0.5 MDL, or MDL will affect estimates of the mean, median, and variance of the observations. In cases where censored data constitute greater than 25% of the sample size, the selection of an appropriate method becomes more arbitrary (Berthouex and Brown 2002).

Model Calibration

After data collection and model setup, parameter values need to be set to accurately predict site-specific responses. Parameters are often established on the basis of the modeler's experience from applications in other systems, trial and error, or with optimization algorithms (Scholten and Refsgaard 2010). This is important because parameters that are set based on measurements (direct or indirect (Barnwell, Brown et al. 2004)) are typically more accurate than those estimated through calibration (Hattermann et al. 2010). While it is recognized that a calibration approach for parameter selection can be problematic (Guadagnini and Neuman 1999), improving data shortfalls can reduce cases where multiple parameter combinations produce the same water-quality predictions (equifinality) (Ebel and Loague 2006).

Model calibration should begin by establishing that certain constituents are predicted correctly before moving onto the more interconnected mechanisms associated with nutrient cycling. First, the flow balance and hydraulics should be verified so that the representation of the residence time and volumes are appropriate. Predicted discharge at downstream locations must match observations. If the values differ substantially, it could be due to inflows or outflows from unknown sources or from groundwater exchanges. These types of sources can be identified using simple differential gauging methods (Ruehl et al. 2006) to provide net changes in flow at various locations throughout a reach. The resulting gains and losses can be assumed to be a distributed groundwater source or abstraction. Abrupt changes in the longitudinal profile of specific conductance from upstream conditions can also provide supporting evidence of the presence of groundwater inflow (Cirpka et al. 2007; Vogt et al. 2010). Once the locations of inflows are roughly identified, the corresponding constituent concentrations require estimation or measurement and can be obtained from nearby seeps (groundwater that surfaces prior to the stream) or shallow groundwater observation wells (Covino and McGlynn 2007; Harvey et al. 1996).

Hydraulic geometry may be specified using either exponential rating curves of velocity (*U*) and depth *(Y)* versus flow *(Q)* (e.g., $Y = a\,Q^b$ *and* $U = c\,Q^d$), or Manning's equation. The coefficients (a and c) and exponents (c and d) for rating curves can be estimated from either long-term gauging station records or hydraulic models (e.g., HEC-RAS). To ensure appropriate flow routing, travel times must be validated. Travel times within the study reach are dependent on hydraulic geometry. If Manning's is used, then Manning's n is typically adjusted to calibrate the depth and velocity since width, channel slope, and side slope should be measured or a rectangular channel assumed. Tracer studies can be helpful in providing data to estimate travel times within the study reach, which in turn, can help gauge the accuracy of average estimates of bottom width, bottom slope, and side slope values.

Longitudinal and diel temperature predictions at different sub-reaches can primarily be adjusted through topographic and riparian shading estimates. Necessary shading information can be estimated using various methods [i.e., SHADE model (Chen 1996)] at each reach element by designating the nearest topographical feature (north, east, and west coordinates and % inclination), vegetation type, and the distance from stream center to the edge of the riparian zone. These types of tools can be used to estimate the hourly percent shading values required by QUAL2Kw. Another consideration is the accuracy of the predicted hydraulic geometry because water temperature response to variations in surface heat fluxes are very sensitive to geometry. At times, it may be necessary to revisit the channel geometry estimates to ensure the accuracy of temperature predictions.

Next, inorganic suspended solids (ISS) settling rate regulates the amount of suspended sediments in the water column, which is important for simulating light penetration through the water column. It can be set directly by adjusting the settling rate to calibrate the ISS predictions to observed stream conditions.

Finally, k_a and SOD can be estimated prior to calibration using various "open-water" methods of determining ecosystem metabolism. A general approach to using ecosystem metabolism methods follows that a change in oxygen over time (*dO/dt*) is a result of oxygen sources (primary production and reaeration) and oxygen sinks (autotrophic and heterotrophic respiration, BOD, and other oxygen consuming reactions within the water column and sediments); however, the relationships describing the change in oxygen is often reduced to Eq. 1:

$$\frac{dO}{dt} = k_a D + GPP - ER \qquad \textbf{(8)}$$

where k_a = stream reaeration rate (d^{-1}), D = DO deficit ($O_{sat} - O$) (mg L^{-1}), *GPP* = gross primary production (mg O_2 L^{-1} d^{-1}), and *ER* = ecosystem respiration (mg O_2 L^{-1} d^{-1}).

Some examples of using this relationship in a stream metabolism context have been established with the Delta Method (Chapra and Di Toro 1991; McBride and Chapra 2005), Nighttime Regression Method (Young et al. 2004), and Inverse Method (Holtgrieve et al. 2010) which simultaneously estimate k_a, *GPP*, and *ER* from the diurnal signal of DO from a single stream station (Eq. 1). With these values established at different points within a study reach, assuming k_a approximations are reasonable, these values can be used to determine which of the widely recognized reaeration formulas provided within QUAL2Kw may be appropriate to predict and represent reaeration under different flow conditions. If diel data are collected from several different river flow conditions then it is possible to derive a site-specific equation to estimate k_a from velocity and depth (e.g., solve for a, b, and c in an equation of the form $k_a = aU^bY^c$). Otherwise, we suggest running the model using each reaeration formula and comparing predictions against the point estimates of k_a from stream metabolism methods. The most appropriate formula within QUAL2Kw can then be selected based on a combination of criteria (e.g., lowest root mean square error, RMSE), appropriateness of formula restrictions or assumptions).

QUAL2Kw has the functionality to estimate SOD based on a sediment diagenesis algorithm (Di Toro, Paquin et al. 1991; Di Toro and Fitzpatrick 1993; Di Toro 2001), but there is often more SOD present than is predicted due to the deposition of organic matter before the time period of model simulation (i.e., during snowmelt runoff) and the deposition of coarse particulate organic matter (CPOM) that typically is not captured by standard sampling techniques. Beyond sediment diagenesis, an additional amount of SOD can be prescribed within the modeling framework but it is handled as a direct sink of oxygen. Since site-specific or reach-integrated SOD measurements are generally not available, there is a need to approximate a reasonable reach scale SOD for each study site. An approach to estimating a maximum SOD is by subtracting *GPP* from *ER*, as defined in Eq. 1. This requires the assumption of autotrophic respiration approximately equaling *GPP* [it may need to be some fraction of *GPP* (Jones et al. 1997)] and any extra oxygen consumption is due to heterotrophic respiration and other oxygen consuming reactions within the sediments and water column. Based on this assumption, a positive value (meaning *ER* is higher than primary production) provides an estimate of total SOD (heterotrophic respiration + oxygen demanding reactions within the sediments) and some oxygen demanding reactions within the water column (e.g., BOD and nitrification). Within QUAL2Kw, it can be assumed that this total SOD would provide a maximum value that includes the prescribed SOD plus the SOD estimated within the sediment diagenesis algorithm [described within (Pelletier and Chapra 2008)]. The assumption that *ER* minus *GPP* equals SOD is assumed appropriate in typical effluent-dominated streams that are relatively shallow and where sediment processes significantly influence the water column DO response. In some situations it is possible that other processes have a more dominant influence on the water column oxygen responses (e.g., chemical reactions within the water column) and these approaches may not be applicable or include more error due to the aforementioned assumptions. Where appropriate, these SOD estimates can provide an upper bound to be used in calibration or an average reach value of SOD could be established and set before auto-calibration.

With a number of parameters set either from direct/indirect measurements or based on these prior manual calibration steps, the remaining parameters that are appropriate to include in model calibration can be auto-calibrated. Using the PIKAIA genetic algorithm (Charbonneau and Knapp 1995) within QUAL2Kw, the number of model runs over which to perform the optimization of the parameter set can be selected (# Model Runs = # of Populations × # of Generations). The parameters that are commonly included in auto-calibration as well as some appropriate parameter ranges are identified in SI Table 2. Within the auto-calibration, a fitness statistic is evaluated for desired state variables as the reciprocal of a weighted average of the normalized RMSE (Pelletier et al. 2006). This tool allows the coefficient of variation of the RMSE between each constituent (model results versus observed data) along with individual weighting factors, to be summarized in a single value that the genetic algorithm seeks to maximize by adjusting all desired parameters.

Case Study

The data collection, model setup, and calibration strategies presented above provide a general framework that can be adapted for different applications of QUAL2Kw or similar models. Once developed and refined, the finalized data collection and modeling approach was applied to a reach in Silver Creek, Utah during low-flow conditions in order to test the validity of the minimal data collection proposed. Assuming constant flow conditions, the data collection prescribed in the generalized approach was used to set up and calibrate the model, with additional data (e.g., SOD measurements and bottom algae samples) collected to evaluate model performance under such conditions.

Silver Creek is a small tributary to the Weber River with land use comprised mainly of Park City, two ski resorts and grazing. The study reach is located 6 miles north of Park City and is approximately 2 km in length near the middle of a 103 km^2 watershed (Figure 2). The climate is typical for high elevation, western mountainous regions, with the majority of the annual precipitation load attributed to winter snowfall and subsequent spring runoff (Whitehead and Judd 2004). During the summer months, some or all of the flow in Silver Creek is diverted upstream of the study reach for irrigation and stock watering purposes, therefore, Silver Creek becomes highly effluent-dependent downstream of the WRF.

Data Collection

The Silver Creek WRF is the major point source for nutrient loading to the Silver Creek study reach although various surface and groundwater seeps also contribute loads (Figure 2 shown as S1, S2, and S3). A small tributary, T1 (Figure 2b), also enters the stream reach one km downstream of the headwater (Station 1). The distance between Stations 2 and 3 was determined according to guidelines set by Grace and Imberger (2006) which designate station spacing based on reaeration estimates derived from depth and velocity measurements of the stream. To verify in-stream processes, intermediate measurement stations were established to provide more detailed hydraulic and water-quality data and are labeled USGS, I1, and I2 at river kilometers 1.8, 1.1 and 0.8, respectively. In addition, a time of travel study between Stations 2 and 3 was conducted using salt (NaCl) as a tracer. Travel time had to be estimated from mean velocity values between Stations 1 and 2 due to low channel flows and large pools, causing a tracer response to become indeterminable at Station 2.

In situ multi-parameter data loggers (YSI 6690 V2, Yellow Springs Instrument Company, Yellow Springs, OH) were deployed from August 22–30, 2011 at Stations 1, 2, T1, and 3 (Figure 2) to collect continuous diel data (five minute intervals) for DO, DO saturation, temperature, pH, conductivity, and chlorophyll *a*. Intermediate stations USGS, I1, and I2 logged from 5 to 15 minute intervals for DO, temperature, conductivity, and pH. YSI protocols were followed for sensor calibration that included a pre-deployment check with all sensors logging in ambient water for 30 minutes prior to deployment and a post-deployment check conducted in the same manner.

Chemistry and nutrient grab samples were collected at Stations 1, 2, T1, and 3 taken twice a day over a two day sampling period per the minimum data requirements protocol. Sampling times for the nutrient and chemistry data were chosen, without prior detailed sampling, to represent the assumed diel variation of constituents governed by the photoperiod with a dawn sample the first day and an afternoon sample the second day (Chapra 2003). Surface seeps S1, S2, and S3 were sampled once during the entire period.

The grab samples were collected according to operating procedures developed by the Utah Division of Water Quality (UDEQ 2012a). They were analyzed for $sCBOD_5$, total nitrogen, total dissolved nitrogen, ammonium, nitrate + nitrite, total phosphorus, total dissolved phosphorus, soluble-reactive phosphorus, chlorophyll *a*, pH, alkalinity, total suspended solids, and volatile suspended solids according to standard methods. From these measured constituents, others were calculated including organic nitrogen, organic phosphorus, detritus and inorganic suspended solids (SI Table 1). Additional samples were collected from the effluent of the WRF to provide an estimate of CBOD oxidation rates. These rates were estimated following EPA method 405.1 (including a nitrification inhibitor) by measuring DO in six reactors each day for 30 days to obtain an average CBOD oxidation rate of 0.103 d^{-1}. Finally, bottom algae estimates from a study conducted in September 2011 produced areal estimates at Stations 1, I2, and 3 and were used to check calibration performance. Values provided are a total of the scaled dry mass that considers the fractional cover of various channel and habitat types for each study section. Macrophyte coverage was omitted from the reported values.

In addition to water-quality samples, supplementary data were collected consisting of geometric, hydraulic, meteorological, and shading information. Width, depth, velocity, and flow measurements were taken several times at each station (S1, S2, T1, and S3) and intermediate stations (I1, I2) along with the high frequency flow record available from the USGS station (USGS 10129900, Silver Creek near Silver Creek Junction, UT). Due to large uncertainties in the flow data, a flow balance study was used to quantify the sources and sinks beyond the major observable inflows in the reach. High variability in channel geometry at Station 1 resulted in inaccurate estimates and required flow to be back-calculated from high frequency flow records from the USGS station and WRF ($Q_{HW} = Q_{USGS} – Q_{WRF}$). Additional distributed inflows and abstractions were quantified based on a channel water balance conducted using differential gauging methods and accounting for known inflows. Meteorological information was downloaded from two weather stations within 25 miles of Silver Creek (National Weather Service stations UTSVC and UTQRY). Data from these stations were used to provide hourly air temperature, dew point temperature, solar radiation, and wind speed during the study period. Longitudinal shading was derived from the SHADE model (Chen 1996) using site-specific vegetation coverage and topographic data extracted from Google Earth™ mapping service.

Model Setup

After collecting the necessary water-quality and site-specific data, model setup ensured water-quality and quantity data was apportioned correctly within the model framework (Table 1). The reach was segmented at 20, 100-meter sections spanning the 2 kilometer study area. Depth, velocity, bottom slope, and side slope point measurements were then interpolated between each reach segment. Each point and distributed source was assigned an average flow value.

Populating headwater data consisted of linear interpolation between available points to estimate hourly values. Point source information used either sonde information to produce a corresponding sine curve, or daily samples (average of two samples) to produce a constant daily concentration. Finally, any water-quality measurements from surface seeps or shallow groundwater observation wells were assigned to either a point (seep) or distributed inflow based on the evidence of groundwater inflow from the flow balance study.

Model Calibration

The methods described within the general calibration approach described previously were followed. However, additional steps included using the CBOD oxidation rates established from the WRF samples to convert the $cBOD_5$ samples to $CBOD_{ultimate}$. The most appropriate reaeration formula and a reach-wide SOD value were estimated for the site using stream metabolism methods estimated specifically using the Inverse Method. An SOD measurement using the chamber method (Hickey 1998) was later collected within the summer season of 2012 near Station 3 and used to evaluate the our approach and the performance. The remaining parameters were then optimized by the auto-calibration algorithm using the fitness statistic by combining the weighted normalized RMSE for each paired observation and prediction at Station 3 for all measured constituents (SI Table 3). Higher weights were assigned to overall indicator constituents such as DO min/max values and pH as well as key constituents such as inorganic phosphorus and nitrogen. While additional water-quality data were available at other stations and could have been incorporated in the fitness statistic for auto-calibration, we opted to rely on a downstream single station as the calibration target per the minimum requirements of the general calibration strategy. The additional data were used to gage the accuracy of the calibration based on limited data.

As described above, similar data collection methods were applied to six additional study reaches located throughout central to northern Utah (SI Figure 1) in order to develop and refine the generalized approach. Generally following the protocols outlined above, these models were populated and calibrated using data only from the most downstream sampling location (even when intermediate data were available). These model results provide further insight regarding the applicability and limitations of the outlined approach to a variety of point source influenced study reaches.

Results

Hourly averaged sonde data for Stations 1-3 and T1 (Figure 3) highlight the daily, longitudinal differences between stations. There are noticeable differences between DO at Station 1 versus those observed at Station 3. The point sources (Stations 2 and T1) differ significantly from each other, most notably between the diel signals, with the WRF (Station 2) experiencing minimal temperature diel variability and high variability for DO while the trends appear reversed for Station T1. Further, the differences of specific conductance between the main channel of Silver Creek and the main tributary would indicate that the source waters are distinct. Also shown are the chlorophyll *a* values (plotted on a log scale) which reflect some reasonable average values, however, the variability within the sonde measurements (as shown by the box plots) illustrates potential drawbacks of relying solely on optical in-situ measurements. Finally, the chemistry and nutrient data results were averaged by station and constituent (SI Table 4). Comparing values between Station 1 and 3 illustrate the effect Stations 2 and T1 have on the downstream concentrations. Also evident is the trade-off between *in situ* sonde measurements of chlorophyll *a* (diel response, larger variability, Figure 3) and the results from the laboratory analysis (fewer measurements, less variability and much lower values).

The flow records taken from 8/18 to 8/30/2011 were averaged together by station and are compared with the net gains and losses results (bar chart) as shown in SI Figure 2. Shown along with the mean daily flow record is the number of points available to calculate the daily average (SI Figure 2b). Even with the low number of records generally available and the uncertainty in measurements, an estimation of the net water balance was necessary to ensure the correct volumes of water were represented in the model. To estimate net gains/losses, the measured mean daily flow value of an upstream station was subtracted from the nearest downstream station (SI Figure 2a). Surface seeps were independently estimated and assigned within the model as distributed sources centered near their surface location which combined represented 0.015 m^3/s or 13% of total stream flow.

Information required from the nutrient, chemistry, and sonde samples illustrate the differences in concentrations between longitudinal stations over a short (two day) period (SI Figure 3). Due to the high variability of the chlorophyll *a* data from the sondes, only results from the laboratory analysis were used for model setup. Further, since specific conductance and pH were measured using both methods (sonde and laboratory analysis) and both types of measurements compared reasonably well, all data were averaged together. All other constituents were summarized as average hourly or daily (in cases of limited samples) for the headwater, point sources, and other intermediate stations. Finally, any samples reported as below MDL were set to 0.5 MDL due to limited sample numbers.

Comparing predicted k_a values to those estimated using the Inverse Method indicated that either the USGS (channel-control/pool-riffle) or the Tsivoglou-Neal methods of reaeration align sufficiently to the data (SI Figure 4). Stream metabolism results (SI Table 5) from Eq. 1 produced GPP and ER values to estimate SOD ranging from 2.1 to 8.6 g $O_2/m^2/d$ with an average value of 5 g $O_2/m^2/d$. Since this average estimate represents the maximum SOD value possible, a slightly lower value of 3 g $O_2/m^2/d$ was prescribed as a conservative starting point for SOD along the entire reach. Due to the number and variability of point estimates, varying SOD by model reach could also be appropriate.

The overall calibration procedure produced reasonable results between observed and predicted values for many of the critical constituents required for accurate representation of stream water quality (Figure 4). Some constituents that matched observations well include inorganic and organic phosphorus as well as ammonium concentrations. Some areas of concern include temperature predictions which miss the mean and minimum observed values at some stations. Some possible explanations could be due to complex groundwater influences or topography and channel incision having a greater effect on shading than captured within the model. The average DO predictions match well, but the minimum and maximum values do not capture the observed diel swings in the upper half of the reach. The most significant source of nitrogen loading is from the WRF in the form of nitrate with values two orders of magnitude less for ammonium and one order less for organic nitrogen. The observed organic nitrogen values at Station 3 were well above the predicted concentrations, likely a legacy of using the differencing method for to derive its values. Predicted bottom algae concentrations were reasonable based on observations during September of 2011 where values ranged from 99 mgA/m^2 for a 70 m^2 section near Station 1, 150 mgA/m^2 for a 120 m^2 section near Station I2, and 226 mgA/m^2 for a 105 m^2 section near Station 3.

Overall, given the minimal amount of data used for model calibration (Station 3 water-quality), it appears that the model represents the observed conditions reasonably well, the exceptions are poor estimates of organic nitrogen, temperature, and DO at the intermediate stations. Including the data from these intermediate stations in the calibration likely would have improved the model calibration.

The six additional study sites throughout Utah (SI Figure 1) all had site specific conditions that influenced sampling, model setup, and/or model calibration (as described within the SI). Regardless, model performance was generally quite good for many key water quality constituents (SI Figure 5-10). It is clear, however, that additional information is necessary to corroborate predictions and that site specific issues influencing data and/or model predictions need to be investigated further. Neilson et al. (2012) provides generalized conclusions and recommendations from this effort.

Discussion

The generalized and basic data collection approach presented within this paper outlines the fundamental data types necessary to set up and conduct a preliminary calibration of the one-dimensional in-stream water-quality model QUAL2Kw assuming steady flow conditions. Since the approach was designed to be applied in the context of addressing various surface water-quality management objectives (e.g., WLAs, TMDLs, NNC), the need to adapt the approach to the case-specific requirements became apparent. The application of this guidance, presented in the context of a detailed case study and prior application to 6 other study areas during development of the generalized approach (see SI), illustrated the utility of both required and potential supplementary data to achieve acceptable predictions. Further, it highlighted the shortcomings of trying to develop low cost, minimalist data collection methods to support model setup and calibration.

Some of the site specific considerations for applying this approach included spatial and temporal sampling needs. Given sufficient background information about a site, some guidance suggests the selection of sampling locations along the reach be placed near the minimum of the DO sag since this will be the area where water-quality standards are likely to be violated (US EPA 1986). In the Silver Creek case study, we adequately captured the effects of the DO sag and nearly captured the downstream location where DO concentrations had returned to the upstream (pre-WRF) concentrations. Unfortunately, data from intermediate sampling locations did not capture this maximum sag in Silver Creek, but were still found useful in identifying locations where the model was obviously not capturing all important processes (Figure 4).

Similar to understanding station locations, bottom algae concentrations are a key factor in many shallow streams (Flynn and Suplee 2011) including the Silver Creek case study. However, there are still limitations when sampling one or two portions of a stream and generalizing the result to an entire study reach. Some methods have been developed which address proper sampling protocols and extrapolating results (CEN 2003), although inevitably, the underlying basis for small sample sizes generally involves time and cost constraints. Additional guidance is needed to derive reliable observations of reach-integrated bottom algae concentrations, to characterize filamentous algae and macrophytes, and how these can be incorporated into water-quality modeling efforts.

Beyond spatial considerations and based on the variability of flow and water-quality, sampling frequency requires more attention for most data types. Despite a low-flow, steady-state assumption, some observed within-day variability can be as large as changes seen on an annual timeframe (Nimick et al. 2011). This is possibly the case for Station 2, TN and TP (SI Figure 3). However, if sufficient diel information of each constituent for a specific site is gathered with the goal to identify a reduced sampling strategy without sacrificing the true signal of the data, the minimal temporal resolution necessary can be determined using spectral analysis with the selection of a sampling interval corresponding to the Nyquist frequency (US EPA 1982). Unfortunately, this data intensive exercise is not feasible for the majority of water-quality assessments. Alternatively, diel variations can be anticipated from previous efforts and daily sampling can be focused on times representing mean concentrations (Harrison et al. 2005; Nimick et al. 2011). The caveat to this is when the variability and magnitude of loading is more significant than the background variability and flow of the stream. In this case, a more extensive survey of water-quality will be required at the loading sources, particularly for unnatural loading signals (independent of photoperiod) commonly seen at a WRF effluent.

In this case study, a daily sampling strategy aimed at capturing anticipated minimum and maximum values appeared to be adequate for many of the necessary constituents. However, complications associated with small sample sizes were compounded when other constituents were estimated (e.g., organic nitrogen, detritus) or when irregular loads from the WRF influenced concentrations of a particular constituent. Further, constituents measured below the detection limit caused a significant bias in model setup and calibration. Due to the limited number of samples taken, the selection of an appropriate method to handle censored values was restricted because they represented more than 25% of the sample size (Berthouex and Brown 2002). The selection of appropriate methods to handle these data became simplified and arbitrary.

Overall prediction uncertainty was created by temporal and spatial data restrictions, difficulty estimating mean daily flows due to highly variable WRF loads, and limited methods for quantifying groundwater exchanges in dynamic systems. These influences were most apparent in the in-stream temperature predictions (Figure 4), but likely influenced the predictions of other constituents. Another key factor was the short reach length and travel time that influenced the ability to identify appropriate parameters that may be key in accurate scenario generation or extended model applications. Further, the assignment of a sine curve to represent the diurnal variation of point sources appears to be an inadequate representation due to an irregular (non-sinusoidal) daily signal (Figure 3). In these cases, the latest version of QUAL2Kw allows for input of hourly values for all sources.

While model selection choices are generally dependent on the management questions and accuracy requirements, the Silver Creek case study provides an example of a situation when the assumption of steady flow with variable concentration may be inappropriate. The QUAL2Kw modeling framework is capable of simulating non-steady flow conditions, but the approach outlined here applies to steady flow conditions due to project budgets often not being sufficient to support data collection for dynamic modeling. Clearly a balance must be achieved between limited data and adequate representation of the key processes and signals, but expanding the data collection to support calibration of non-steady conditions may avoid problems in systems like Silver Creek that experience drastic changes in loading from day to day. Kannel and colleagues (2011) mention similar limitations and emphasize that despite these challenges and depending on the system, the time and cost advantages of assuming steady flows may outweigh the additional cost associated with calibration of continuous simulation of non-steady flows. In an effort to develop simple methods to set two key parameters, k_a and SOD, using DO time series and open-water metabolism methods, it appears that the proposed approaches are reasonable for Silver Creek. The assumptions associated with ER minus GPP being equivalent to a maximum SOD will not be applicable to all systems; however, in some circumstances, the ability to take advantage of already existing data to complete these calculations appears acceptable. Since there is no clear consensus on which methods are most appropriate for measuring or estimating SOD (Viollier et al. 2003) due to temperature gradients (Otubu et al. 2006), velocity dynamics (Nakamura and Stefan 1994), and spatial heterogeneity (Mugler et al. 2012), we chose to test our estimates by comparing them with SOD measurements from *in situ* chambers deployed in 2012. The chambers produced values near 3 g O_2/m^2/day while our values ranged from 2.1–8.6 g O_2/m^2/day with an average of 5.2 g O_2/m^2/day based on the ER/GPP differencing calculations. These similarities suggest that the differencing approach is a reasonable way to set or bound SOD values before or during model calibration.

Reaeration is notoriously difficult to estimate (Genereux and Hemond 1992) though direct measurement techniques using a tracer such as propane to estimate the gas exchange coefficient are superior to deriving k_a using physical characteristics of bottom slope, water depth, and stream velocity (McCutchan et al. 1998). Using stream metabolism methods in the context of k_a values is not new (Odum 1956); however, using these values at many locations to inform the selection of the internal model formula has shown to be potentially useful. Unfortunately, in the case of Silver Creek, two formulas were found to have the best RMSE values although they differed significantly from one another (one higher, one lower). More importantly, neither were a good fit to the data due to simplified hydraulics within the model. While a poor formula fit might inspire a modeler to set a reach-wide k_a value using the average of observed data points, this practice can be problematic for scenarios run under different flow conditions. An alternate approach could be to calibrate the model based on the observed k_a values and use the data to select the most appropriate formula for subsequent scenarios.

Most process-oriented models are under-determined, wherein there exist more parameters than state variables to define them (Reckhow and Chapra 1999). Although a sensitivity analysis can help to identify the key processes influencing the state variables, there is an obvious need to decrease the number of parameters and potentially come up with narrower ranges to confine auto-calibration estimates. The potential number of calibration parameters is extremely high in QUAL2Kw, but without more information regarding which parameters are unimportant, it is not clear which should be dropped from the auto-calibration. More effort is needed to identify the most sensitive parameters of a system, narrow the reasonable ranges of those parameters, and set those parameter rates according to appropriate site-specific conditions. It is also important to identify which outputs must be included in a fitness statistic since the objective function (i.e., fitness) guides the auto-calibration. Given what we know of receiving streams downstream of WRFs, SOD and bottom plant growth are primary factors in governing DO dynamics (Chapra 2008; Utley et al. 2008) and future efforts should focus on developing appropriate sampling or simplified modeling approaches to represent these processes.

This study presents a minimalist approach to model setup and calibration that is reproducible and applicable to a diverse set of water-quality modeling problems. Applying this approach to the Silver Creek case study resulted in reasonable model predictions that captured many of the dominant processes that affect DO. However, for most problems this approach would only provide an initial framework for preliminary data collection that would be adapted as needed. Additionally, a sensitivity analysis, additional model calibration, and model corroboration or validation would have to occur before applying it to management decisions. Regardless, this basic data collection approach results in a judicious use of resources while assisting in identifying the key factors requiring additional investigation.

Conclusion

In this paper, we developed a general data collection methodology to support QUAL2Kw model setup, explored methods for estimating key model parameters, and addressed the effectiveness of these methods with a case study of an effluent-dominated stream system. To minimize data collection costs, we identified the nominal number of sampling locations and minimal required data types for a WRF-dominated system. In the context of a case study we illustrated the utility of collecting grab samples over a two day period with one sample collected at dawn and the other at dusk. These data were supplemented with *in situ* sonde information to capture the daily variability of other constituents governed by the photoperiod. We found that this basic approach provided adequate information for model setup and calibration and reasonable predictions for the Silver Creek case study. However, we recognize that other sampling frequencies may be necessary for other study sites and objectives.

Some challenges were identified as we translated the collected data into model setup including estimation of correct flow rates and volumes, designation of appropriate loading values due to variable point source loads, and determination of appropriate calibration/fitness endpoints. The identification of key parameters (k_a and SOD) using data collected within the proposed methodology provided a means to decrease model uncertainty by reducing model parameters being calibrated. Future work to reduce the ranges of model calibration parameters, identification of sensitive parameters, and/or development of additional methods to set more model parameters based on site-specific conditions will help to increase confidence in model predictions. While this approach has merit as a starting point for WLAs, TMDLs, and in helping develop nutrient criteria, it should not be used as a "one-size-fits-all" strategy, but rather incorporated in an adaptive management strategy to guide more appropriate site-specific data collection schemes that will facilitate predictions needed to address management objectives.

References

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

Bartholow, J. M. 1989. Stream temperature investigations: field and analytic methods. Instream Flow Information Paper No. 13. U.S. Fish and Wildlife Service. Biological Report 89(17)

Systems Ecology. Covino, T. P., and B. L. McGlynn. 2007. Stream gains and losses across a mountain-to-valley transition: impacts on watershed hydrology and stream water chemistry. Water Resources Research 43(10):1–14.

Di Toro, D. M. 2001. Sediment Flux Modeling. Wiley-Interscience, New York, New York.

Di Toro, D. M., and J. F. Fitzpatrick. 1993. Chesapeake Bay sediment flux model. Waterways Experiment Station, U.S. Army Corps of Engineers. Vicksburg, Mississippi. Report EL-93-2:316.

Di Toro, D. M., P. R. Paquin, et al. 1991. Sediment Oxygen Demand Model: Methane and Ammonia Oxidation. Journal of Environmental Engineering 116(5):945–986.

Dunnette, D. A. 1980. Sampling frequency optimization using a water quality index. Water Pollution Control Federation 52(11):2807–2811.

Ebel, B. A., and K. Loague. 2006. Physics-based hydrologic-response simulation: seeing through the fog of equifinality. Hydrological Processes 20:2887–2900.

Ecology and Washington State Department of Ecology. 2003. Shade.xls—a tool for estimating shade from riparian vegetation.

Ensign, S. H., and M. W. Doyle. 2006. Nutrient spiraling in streams and river networks. Journal of Geophysical Research 111:1–13.

Facchi, A., C. Gandolfi, and M. J. Whelan. 2007. A comparison of river water quality sampling methodologies under highly variable load conditions. Chemosphere 66(4):746–756.

Flynn, K., and M. Suplee. 2011. Using a computer water quality model to derive numeric nutrient criteria: Lower Yellowstone River. Montana Department of Environmental Quality, Helena, Montana.

Fogle, A. W., J. L. Taraba, and J. S. Dinger. 2003. Mass load estimation errors utilizing grab sampling strategies in a karst watershed. Journal of the American Water Resources Association 39(6):1361–1372.

Gammons, C. H., J. N. Babcock, S. R. Parker, and S. R. Poulson. 2011. Diel cycling and stable isotopes of dissolved oxygen, dissolved inorganic carbon, and nitrogenous species in a stream receiving treated municipal sewage. Chemical Geology 283(1):44–55.

Genereux, D. P., and H. F. Hemond. 1992. Determination of gas exchange rate constants for a small stream on Walker Branch Watershed, Tennessee. Water Resources Research 28(9):2365–2374.

Grace, M., and S. Imberger. 2006. Stream metabolism: performing and interpreting measurements. Water Studies Centre Monash University, Murray Darling Basin Commission and New South Wales Department of Environment and Climate Change, 204.

Guadagnini, A., and S. P. Neuman. 1999. Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains. Water Resources Research 35(10):2999–3018.

Gunderson, L., and J. Klang. 2004. Lower Minnesota River dissolved oxygen TMDL. Minnesota Pollution Control Agency.

Harrison, J. A., P. A. Matson, and S. E. Fendorf. 2005. Effects of a diel cycle on nitrogen transformations and greenhouse gas emissions in a eutrophied subtropical stream. Aquatic Science 67:308–315.

Harvey, J. W., B. J. Wagner, and K. E. Bencala. 1996. Evaluating the reliability of the stream tracer approach to characterize stream-subsurface exchange. Water Resources Research 32(8):2441–2451.

Hattermann, F. F., P. Willems, and Z. W. Kundzewicz. 2010. Modelling—a primer for practioners. Pages 55–86 *in* F. F. Hattermann and Z. W. Kundzewicz (editors). Water Framework Directive: Model Supported Implementation, IWA Publishing, London, UK.

Hazelton, C. 1998. Variations between continuous and spot-sampling techniques in monitoring a change in river-water quality. Water and Environment Journal 12(2):124–129.

Henderson-Sellers, B., and A. Henderson-Sellers. 1996. Sensitivity evaluation of environmental models using fractional factorial experimentation. Ecological Modelling 86(2–3):291–295.

Henjum, Mi. B., R. M. Hozalski, and C. R. Wennen. 2010. A comparison of total maximum daily load (TMDL) calculations in urban streams using near real-time and periodic sampling data. Journal of Environmental Monitoring 12(1):234–241.

Hester, E. T., and M. W. Doyle. 2011. Human impacts to river temperature and their effects on biological processes: a quantitative synthesis. Journal of the American Water Resources Association 47(3):571–587.

Hickey, C. W. 1998. Benthic chamber for use in rivers: testing against oxygen mass balances. Journal of Environmental Engineering 114:828–845.

Holtgrieve, G. W., D. E. Schindler, T. A. Branch, and Z. T. A'mar. 2010. Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics. Limnology and Oceanography 55(3):1047–1063.

Jones, J. B., J. D. Schade, S. G. Fisher, and N. B. Grimm. 1997. Organic Matter Dynamics in Sycamore Creek, a Desert Stream in Arizona, USA. Journal of the North American Benthological Society 16(1):78–82.

Kannel, P. R., S. R. Kanel, S. Lee, Y.-S. Lee, and T. Y. Gan. 2011. A review of public domain water quality models for simulating dissolved oxygen in rivers and streams. Environmental Monitoring Assessment 16:183–204.

Lettenmaier, D. P., E. R. Hooper, C. Wagoner, and K. B. Faris. 1991. Trends in stream quality in the continental United States, 1978–1987. Water Resources Research 24(3):327–339.

McBride, G. B., and S. C. Chapra. 2005. Rapid calculation of oxygen in streams: approximate delta method. Journal of Environmental Engineering 131(3):336–342.

McCutchan, J. H., W. M. Lewis, and J. F. Saunders. 1998. Uncertainty in the estimation of stream metabolism from open-channel oxygen concentrations. Journal of the North American Benthological Society 17(2):155–164.

Mugler, C., C. Rabouille, B. Bombled, and P. Montarnal. 2012. Impact of spatial heterogeneities on oxygen consumption in sediments: experimental observations and 2D numerical modeling. Journal of Geochemical Exploration 112:76–83.

Nakamura, Y., and H. G. Stefan. 1994. Effect of flow velocity on sediment oxygen demand: A theory. Journal of Environmental Engineering 120:996–1016.

National Research Council. 2001. Assessing the TMDL Approach to Water Quality Management. The National Academies Press, Washington, DC.

National Research Council. 2007. Models in environmental regulatory decision making. National Research Council, Washington, DC.

Neilson, B. T., and S. C. Chapra. 2003. Integration of water quality monitoring and modeling for TMDL development. Water Resources Impact 5(1):9–11.

Neilson, B. T., A. J. Hobson, N. von Stackelberg, and J. Ostermiller. 2012. Using Qual2K Modeling to Support Nutrient Criteria Development and Wasteload Analyses in Utah. Final Project Report. Utah Department of Environmental Quality. Division of Water Quality. September 2012.

Nimick, D. A., C. H. Gammons, and S. R. Parker. 2011. Diel biogeochemical processes and their effect on the aqueous chemistry of streams: A review. Chemical Geology 283(1):3–17.

Odum, H. T. 1956. Primary production in flowing waters. Limnology and Oceanography 1(2):102–117.

Orlob, G. T. 1992. Water-quality modeling for decision making. Journal of Water Resources Management and Planning 118:295–307.

Ort, C., M. G. Lawrence, J. Reungoat, and J. F. Mueller. 2010. Sampling for PPCPs in wastewater systems: comparison of different sampling modes and optimization strategies. Environmental Science & Technology 44(16):6289–6296.

Otubu, J. E., J. V. Hunter, K. L. Francisco, and C. G. Uchrin. 2006. Temperature effects on tubificid worms and their relation to sediment oxygen demand. Journal of Environmental Science and Health, Part A 41(8):1607–1613.

Pelletier, G. J., and S. C. Chapra. 2008. QUAL2Kw theory and documentation. Washington State Department of Ecology, Olympia, WA.

Pelletier, G. J., S. C. Chapra, and H. Tao. 2006. QUAL2Kw—A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. Environmental Modelling & Software 21(3):419–425.

Reckhow, K. H., and S. C. Chapra. 1999. Modeling excessive nutrient loading in the environment. Environmental Pollution 100:197–207.

Ruehl, C., A. T. Fisher, C. Hatch, M. L. Huertos, G. Stemler, and C. Shennan. 2006. Differential gauging and tracer tests resolve seepage fluxes in a strongly-losing stream. Journal of Hydrology (330):235–248.

Scholten, H., and J. C. Refsgaard. 2010. Quality assurance in model-based water management. Pages 1-26 *in* P. A. Vanrolleghem (editor). Modeling Aspects of Water Framework Directive Implementation, IWA Publishing, London.

Stahl, J. K., and R. L. Smith. 2002. TMDL analysis for Limekiln Brook, Danbury, Connecticut. Department of Environmental Protection, Hartford, CT.

Turner, D. F., G. J. Pelletier, and B. Kasper. 2009. Dissolved Oxygen and pH Modeling of a Periphyton Dominated, Nutrient Enriched River. Journal of Environmental Engineering 135(8):645–652.

UDEQ. 2000. TMDL for East Canyon Creek. Utah Department of Environmental Quality, Salt Lake City, UT.

UDEQ. 2012a. Field data collection for QUAL2Kw model build and calibration, standard operating procedures, Version 1.0. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, UT.

UDEQ. 2012b. Utah wasteload analysis procedure. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah, 1–23.

US EPA. 1982. Handbook for sampling and sample preservation of water and wastewater. United States Environmental Protection Agency, Washington, DC.

US EPA. 1986. Stream sampling for waste load allocation applications. United States Environmental Protection Agency, Washington DC.

US EPA. 1995. Technical guidance manual for developing total maximum daily loads. United States Environmental Protection Agency, Washington, DC.

US EPA. 1997. Streams and Rivers. Technical guidance manual for performing waste load allocations, Streams and Rivers, United States Environmental Protection Agency, Washington, DC.

US EPA. 2000. Nutrient criteria technical guidance manual: rivers and streams. United States Environmental Protection Agency, Washington, DC.

US EPA. 2002a. TMDL for total mercury in fish tissue residue. United States Environmental Protection Agency: Region 4, Lake Oconee, Georgia.

US EPA. 2002b. Guidance on choosing a sampling design for environmental data collection. United States Environmental Protection Agency, Washington, DC.

US EPA. 2009. National Rivers and Streams Assessment Field Operations Manual. United States Environmental Protection Agency, Washington, DC.

Utley, B. C., G. Vellidis, R. Lowrance, and M. C. Smith. 2008. Factors affecting sediment oxygen demand dynamics in blackwater streams of Georgia's coastal plain. Journal of the American Water Resources Association 44(3):751–753.

Viollier, E., C. Rabouille, S. E. Apitz, E. Breuer, G. Chaillou, K. Dedieu, Y. Furukawa, C. Grenz, P. Hall, F. Janssen, J. L. Morford, J. C. Poggiale, S. Roberts, T. Shimmield, M. Taillefert, A. Tengberg, F. Wenzhofer, and U. Witte. 2003. Benthic biogeochemistry: state of the art technologies and guidelines for the future of in situ survey. Journal of Experimental Marine Biology and Ecology 285:5–31.

Vogt, T., E. Hoehn, P. Schneider, A. Freund, M. Schirmer, and O. A. Cirpka. 2010. Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream. Advances in Water Resources 33(11):1296–1308.

von Stackelberg, N. O., and B. T. Neilson. 2014. A collaborative approach to calibration of a riverine water quality model. Journal of Water Resources Planning and Management, 140(3):393–405, doi: 10.1061/(ASCE)WR.1943-5452.0000332.

Whitehead, J., and H. Judd. 2004. Silver Creek total maximum daily load for dissolved zinc and cadmium. Division of Water Quality, Salt Lake City.

Whitehead, P. G., R. L. Wilby, R. W. Battarbee, M. Kernan, and A. J. Wade. 2009. A review of the potential impacts of climate change on surface water quality. Hydrological Sciences 54(1):101–123.

Young, R., C. Townsend, and C. Matthaei. 2004. Functional indicators of river ecosystem health—an interim guide for use in New Zealand. Cawthron Institute, Nelson, New Zealand.

Zhang, J., and C. Zhang. 2012. Sampling and sampling strategies for environmental analysis. International Journal of Environmental Analytical Chemistry 92(4):466–478.

Table 1. General information required for QUAL2Kw model setup.

QUAL2Kw Input	Information Required
Reach	Reach segmentation
	Hydraulic characteristics
	% suitable substrate
	Bottom algae % cover
	Sediment Oxygen Demand (SOD)
	Thermal properties
Initial Conditions	Constituent concentrations[1, 2]
Headwater Data	Average flow
	Hourly mean concentrations[1]
Weather Data	Hourly mean values[3]
Point Sources	Average flow
	Daily mean,[1] Range/2, Time of Max
Distributed Sources	Average flow
	Daily mean concentrations[1]
Point and Distributed Abstractions	Average flow
Rates	Primarily set in calibration
	Literature informed ranges of parameters

1. See SI Table 1 for a list of constituents required by QUAL2Kw.
2. Optional.
3. The model is interpolating between the hourly values for each time step; therefore, the input data should ideally be an average of the instantaneous data on the hour. If using hourly averages, the averages should be centered on the hour.

Table 2. Site characterization data types, procedures and locations.

Data Type	Procedure	Locations	Reasoning
Average Cross Sectional Velocity	Velocity cross-sectional profile obtained from velocity-area method of discharge measurements. Information from HEC-RAS modeling applications can also be extracted to supplement data collected.	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of velocity in different reaches to compare with the predicted velocities. This can be used with the depth and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Cross Sectional Depth	Average depth can be obtained from velocity-area method of discharge or independently estimated cross-sectional depth profiles.	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of depths in different reaches to compare with the predicted depths. This can be used with the velocity and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Channel Bottom Width	Bottom width estimates are calculated using the formula: Top Width (m) $-\text{Depth}_{ave}$ (m) x 1/tan[radians(°SS_{LEW})] $-\text{Depth}_{ave}$ x 1/tan[radians (°SS_{REW})].	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model Input. Top widths, side slopes, and bottom slopes are measured at consistent increments along the channel. From these data, bottom width estimates can be calculated using side slope, average depth, and top width values.
Channel Bottom Slope	Measured with a survey level or clinometers, protocols described within EMAP documentation[1].	Should estimate bottom slope from beginning to end of study reach at each station and/or when changes in bottom slope are observed.	Model Input. Bottom slope affects travel time and can be adjusted along with Manning's n for to achieve proper estimates.
Channel Side Slope	Measured with a clinometer or by visual inspection, protocols described within EMAP documentation[1].	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input. Can be used to calculate bottom width from measured top widths.
Weather Data	Onsite or nearest weather station.	Near study site would be most appropriate and 15–30-minute data are preferred, hourly estimates required.	Wind speed, air temperature, shortwave solar radiation, humidity/dew point temperature are all used within the model as forcing information.
Tracer Study	Inject tracer at Station 1 or 2 and measure response at Station 3. Can also use HEC-RAS model if available.	Measure tracer response at Station 3, but additional locations along the study reach would be beneficial to capture heterogeneity and identify potentially significant groundwater sources.	Provides information regarding average travel time through system and can be used in calibration of hydraulic parameters (e.g., Manning's roughness coefficient).
Substrate Type	Protocols described within EMAP documentation[1]	Information should be gathered at cross sections in sub-reaches that represent the variability in substrate types.	Provides a method to approximate the Manning's roughness coefficient and determine fraction of bottom substrate appropriate for bottom algae.
Shading	Estimated with shading model (e.g., SHADE[2]) to predict effective shade from topography and riparian vegetation	Information should be gathered at locations that represent the variability in shading.	Model input. If riparian or topographic shading drastically influences in-stream temperatures, estimates of the shading % for each hour of a day will be necessary to scale the incoming shortwave solar radiation.

1. (US EPA 2009)
2. (Ecology and Washington State Department of Ecology 2003)

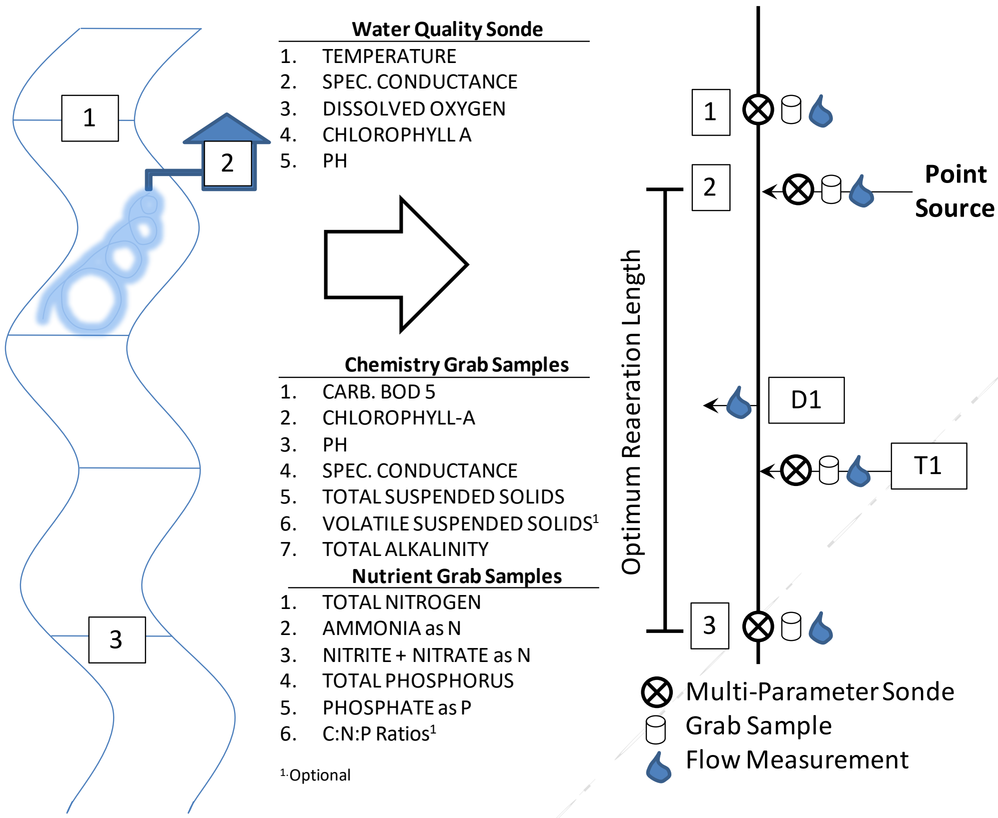


Figure 1. Generalized data collection locations are shown with the required locations of flow measurements, multi-parameter water quality sondes, and chemistry and nutrient samples. Headwater (upstream boundary condition) is designated by Station 1, the primary point source is represented by Station 2, tributaries are denoted with T1, diversions with D1, and the downstream calibration station is shown as Station 3.

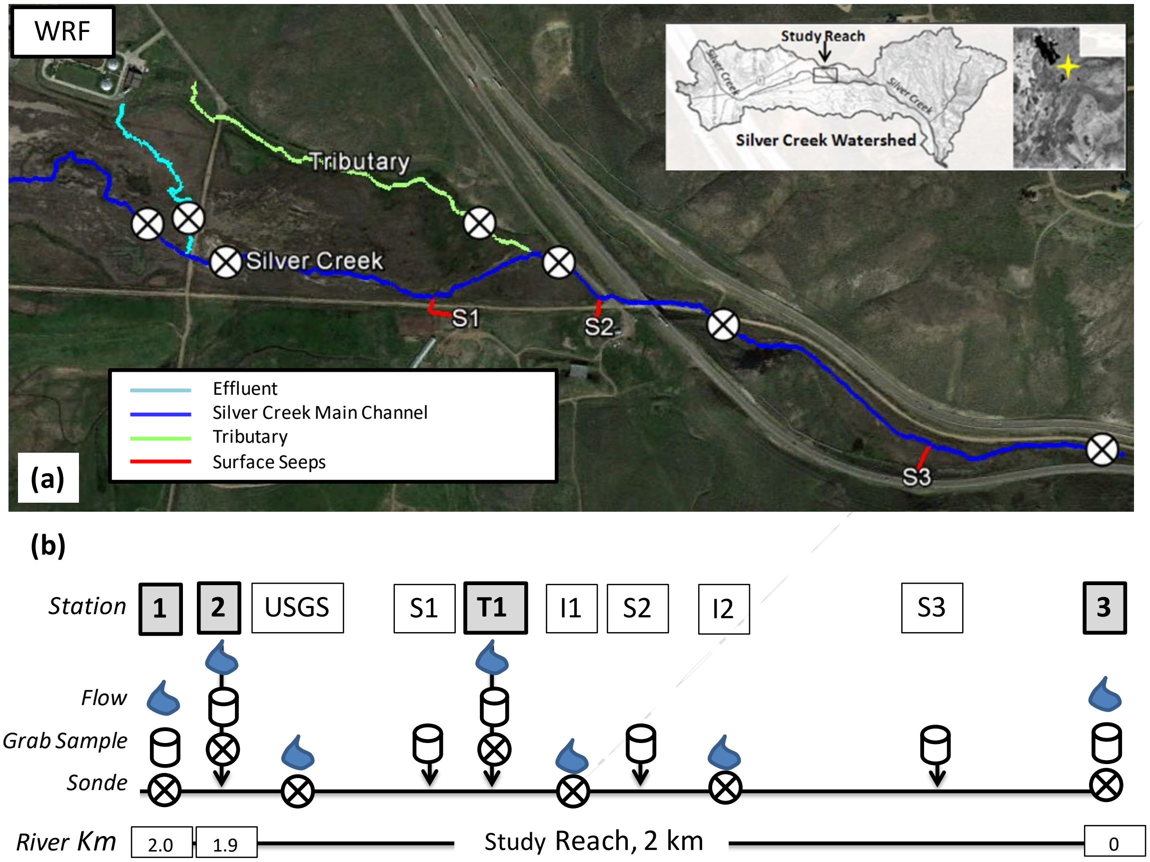


Figure 2. The location of the study area spans just upstream of the discharge of the Silver Creek WRF to Silver Creek and extends downstream approximately 2 km for a total travel time of 0.13 days (≈ 3 hours). Water quality is continuously monitored at a USGS station (river km 1.8) and the Silver Creek WRF (a). The aerial map is simplified into a site schematic for the study reach (b). Major stations include (1) headwater at river km 2.0, (2) WRF point source at river km 1.9, (T1) a tributary to the stream at river km 1.2, and (3) a downstream calibration station at river km 0. Also shown are intermediate measurement stations denoted as I1 and I2 at river kilometers 1.1 and 0.8, respectively. Visible waters flowing to the stream are denoted as surface seeps with S1, S2, and S3 at river kilometers 1.4, 1.0, and 0.3, respectively.

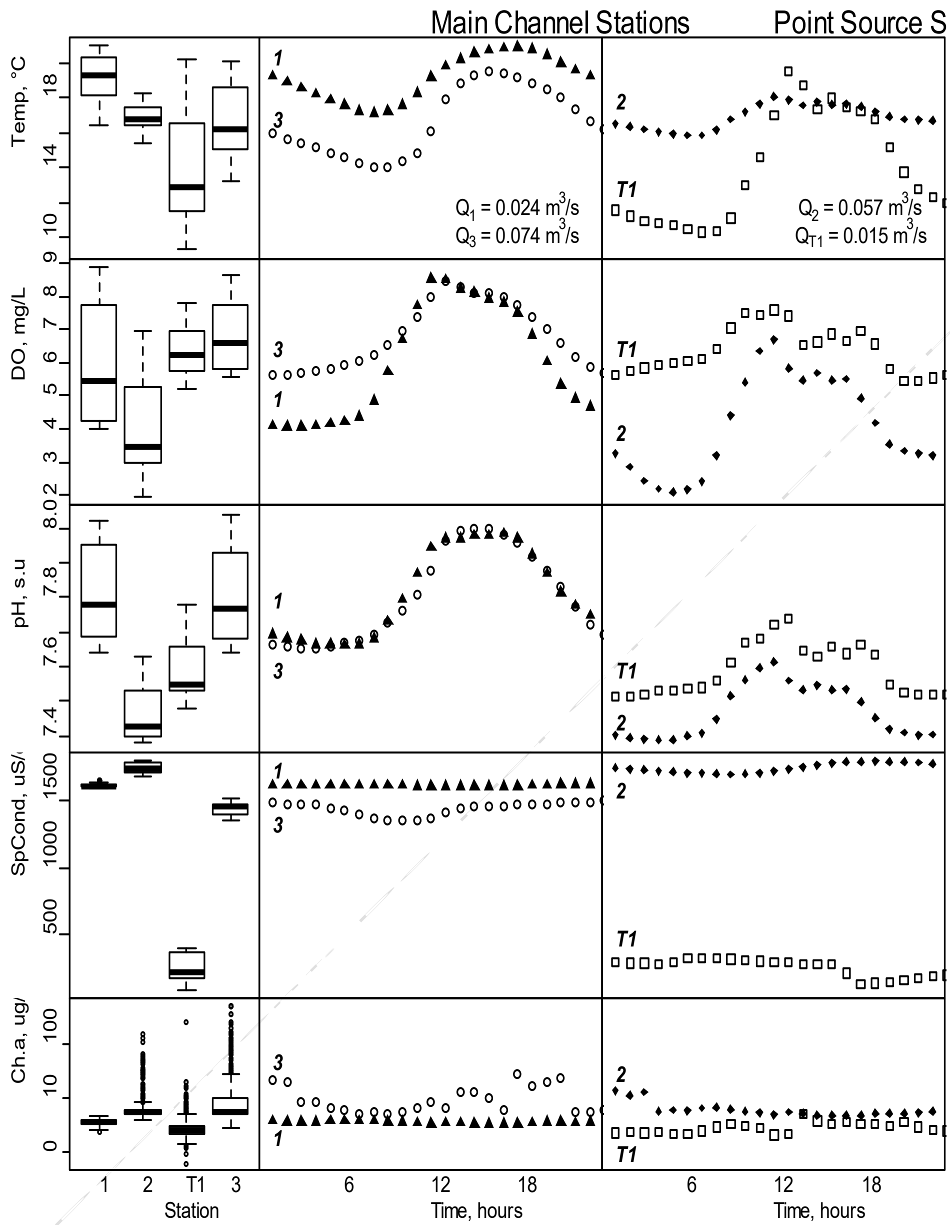


Figure 3. Sonde data collected at Station 1 (headwater), Station 2 (WRF), Station T1 (tributary), and Station 3 (calibration station) measuring temperature, DO, pH, specific conductance, and chlorophyll-*a*. The first column shows each station placed in order longitudinally along the reach. The second column shows the difference from in-stream stations, the headwater and the downstream calibration station. The third column shows the major point sources influencing the stream reach (WRF, T1). Diel values were averaged hourly over two days. Note that chlorophyll-*a* values were log10 transformed due to many extreme values near the maximum detection limit.

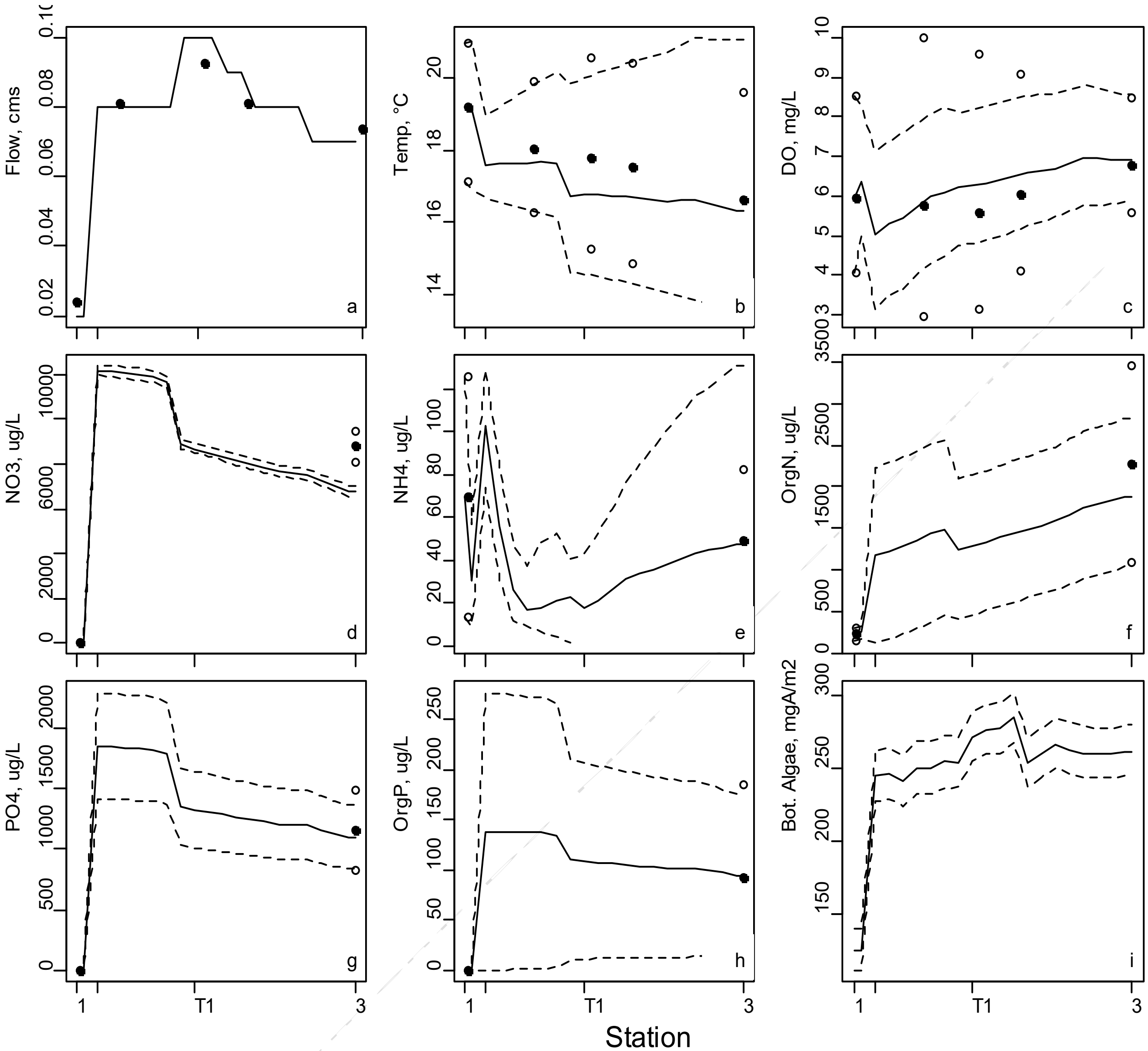


Figure 4. Comparison of predicted versus measured data for a) flow, b) water temperature, c) DO, d) nitrate, e) ammonium, f) organic nitrogen, g) inorganic phosphorus, h) organic phosphorus, and i) bottom algae of Silver Creek (X axis is in river kilometers) for the Qual2Kw model calibration. The solid lines indicate model predictions, dashed lines are minimum and maximum predicted values, solid circles are average daily measurements and white circles are daily minimum and maximum observed concentrations.

Sources of Uncertainty in Nutrient Collection Methods below a Point Source

Hayden Campbell, Undergraduate Researcher (Biology)
Michelle Baker (Biology), Beth Neilson (Civil and Environmental Engineering), and Susan Durham (Ecology Center Statistician).
14 February 2014

Abstract

The goal of this study was to determine what aspects of sampling and sample storage could lead to uncertainty when taking samples below a point source. Sources of uncertainty studied were the locations where the samples were taken to assess if nutrients were adequately mixed within a cross-section, different filtration techniques, dilution errors, analytical uncertainty, and storage time. Bootstrapping analyses were used to determine whether mixing and dilution errors led to uncertainty, while one-way ANOVAs were used to evaluate filtration techniques and storage time. Sample spikes to determine percent recovery of nutrients and repeat sample analyses are routinely performed as part of the lab quality assurance/quality control plan (QA/QC) and are used here to evaluate analytical uncertainty. Comparison of coefficients of variation (COVs) of samples collected within a cross section at four locations, above, at, and below a point source, revealed that mixing of nutrients within a cross section appeared to be different at the different locations. The filtration devices analyzed were an electric pump and a manual syringe. These two devices gave statistically similar results in nitrate and soluble reactive phosphorus concentrations ($p > 0.05$), but syringe-filtered samples had significantly higher ammonium concentrations ($p < 0.05$). Dilution error was determined by comparing seven diluted samples with the original sample with which they were made. Dilutions proved to have the highest uncertainty relative to other treatments. The diluted samples were consistently higher than the original sample for all nutrients and were more variable than lab QA/QC duplicates for ammonium and soluble reactive phosphorus. Analytical uncertainty was found to be less than uncertainty associated with sample collection and storage except for unanticipated protocol failure. For this study, QA/QC data beyond 20% were considered fails, and the samples required reanalysis. In most cases the percent recovery of spiked samples and coefficients of variation of samples repeatedly analyzed were much less than 20%. However, ammonium and total nitrogen incurred the most failures. Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations ($p < 0.05$), however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates.

Introduction

Nutrient samples are often collected below point sources, such as wastewater treatment plants, to ascertain if nutrient quantities exceed in-stream water quality standards. When analyzing these samples, it is imperative that the samples at the time of analysis are representative of the samples at the original time of sampling. In order to ensure that nutrient samples are reliable, this study tested five sources of uncertainty that had the potential to cause unreliable nutrient measurements. These sources included the location within a cross-section where samples were taken in order to assess whether inadequate mixing was occurring within the stream, different filtration techniques, dilution errors, analytical uncertainty, and storage time. Samples were collected in Silver Creek around the Silver Creek Water Reclamation Facility near Park City, Utah.

This research emerged due to recognized anomalies in previous nutrient sampling completed below point sources. Previously, when these analyses were completed, comparisons were made between total phosphorus and constituent phosphorus concentrations (e.g., soluble reactive phosphorus [SRP]). The amount of SRP measured was greater than the amount of total phosphorus measured. Additionally, similar anomalies were found when comparing constituent dissolved nitrogen concentrations and total nitrogen concentrations. This led to the recognition of sampling and/or analytical errors but did not reveal the source of the error. This research was conducted in order to determine what potential sources of uncertainty could have led to these anomalies.

Methods

Sampling Location

Four locations were chosen along Silver Creek where samples were collected to test whether mixing could cause different nutrient concentrations depending on where in the cross section the samples were collected. These locations included one above where the wastewater treatment plant effluent enters the stream (Above WWTP, approximately 13 meters), one in the wastewater effluent (Point Source), and two below where the wastewater treatment plant effluent enters the stream (Below (I) WWTP, approximately 103 meters, and Below (II) WWTP, approximately 717 meters) (Figures 1a and 1b). Seven 1,000 mL grab samples were collected at each of the four cross-sections. Each sample was collected at a different location within the cross section, both at varying distances across the cross section and at varying depths (Figure 2). From each 1,000 mL sample, two sub-samples were taken. One sub-sample was 120 mL that was not filtered, and was used to analyze for total nitrogen (TN) and total phosphorus (TP). The second sub-sample was filtered with a syringe and was analyzed for nitrite+nitrate-N, ammonium-N, and soluble reactive phosphorus (SRP). Samples were collected in the following order to prevent contamination: Above WWTP , Below (II) WWTP, Below (I) WWTP, Point Source.

The coefficients of variation were calculated for each cross section and compared to determine if they were statistically different. This was done in order to identify if the cross sections, whether above or below the wastewater treatment plant, had different mixing patterns. Three parametric tests were run to compare all four coefficients of variation, including the Modified Bennet's test, the Wald Test, and the Modified Miller Test (Jafari and Kazemi, 2013). These tests would determine statistical significance if the calculated p-value was less than 0.05.

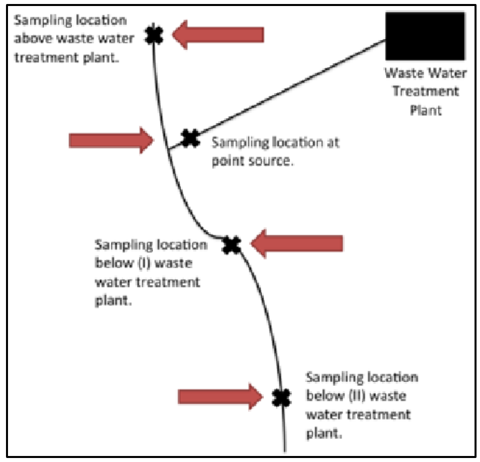


Figure 1a. Locations for testing inadequate mixing.

Figure 1b. Google Earth image of sampling locations

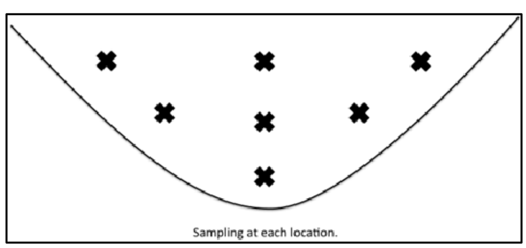


Figure 2. Approximate locations in cross section of stream where grab samples are obtained.

Since these tests compared all four sites together, another statistical test was required in order to determine which sites, between the four, actually were statistically different. A nonparametric bootstrap was used to compare the coefficients of variation between two individual sites to determine which sites yielded different mixing patterns.

Filtration Techniques

The samples collected for analyzing whether differences in nutrient concentrations exist when filtering with a syringe or an electric pump were collected at the farthest location downstream (Below (II) WWTP) (Figure 3). A 5-gallon bucket was collected in the middle of the stream at this location. From this bucket, which was kept mixed with a hand mixer, 12 samples were collected using the electric geopump (Figure 4), and 12 samples were collecting using a manual syringe. Each of these samples was analyzed for ammonium, nitrate, and soluble reactive phosphorus (SRP). One-way ANOVAs were used to compare the average concentrations of each nutrient for both filtering methods.

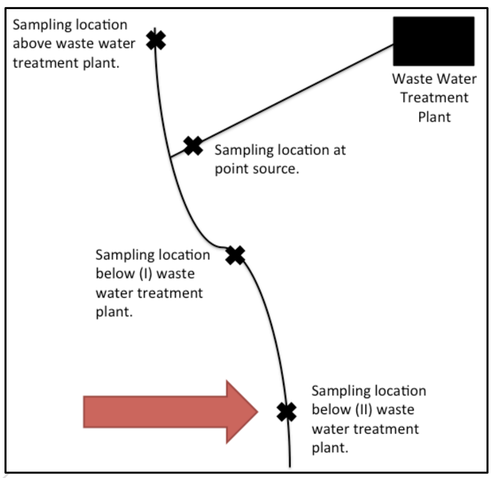


Figure 3. Location of sampling to contrast filtering methods.

Figure 4. Electric pump and syringes used to contrast filtering methods.

Dilution Error

From another five-gallon bucket collected Below (II) WWTP, five 120 mL grab samples were taken and five samples were filtered with a syringe. These samples were repeat samples to determine how variable nutrient concentrations were when samples were collected from the same location. The unfiltered samples were analyzed for total nitrogen and total phosphorus, and the filtered samples were analyzed for ammonium, nitrate, and soluble reactive phosphorus. Two unfiltered samples and two filtered samples were used to make dilutions to determine how variable nutrient concentrations were when making dilutions. Five 1:100 dilutions were made for each of the four samples. The dilutions made on the unfiltered samples were analyzed for total nitrogen and total phosphorus, and the dilutions made on the filtered samples were analyzed for ammonium, nitrate, and soluble reactive phosphorus. These concentrations were compared to the original samples.

Analytical Uncertainty

Lab quality assurance and control (QA/QC) uses results from spiked samples, duplicates, and certified reference materials to assess lab analyses. For this study, QA/QC data beyond 20% were considered fails, and the samples required reanalysis. For each nutrient, the number of fails were tallied for duplicate samples and spiked samples to determine what nutrients incurred the most analytical failure.

Storage Time

Samples to assess the reliability of freezing were collected from Below (II) WWTP. A 4,000 mL grab sample was collected from the middle of the stream and transported back to the lab. Twenty-eight samples were made from this 4,000 mL sample and placed in the freezer. Because these samples were unfiltered, they were analyzed for total nitrogen and total phosphorus. Seven samples were analyzed after one week of freezing, seven samples were analyzed after three weeks of freezing, seven samples were analyzed after six weeks of freezing, and seven samples were analyzed after twelve weeks of freezing. The average concentrations were compared using one-way ANOVAs to test for a statistical significance.

After the seven samples were analyzed after one week, they were placed back in the freezer and analyzed again at three weeks, six weeks, and twelve weeks of freezing. The samples analyzed after three weeks were also placed back in the freezer and analyzed again at six weeks and twelve weeks of freezing. The same was done for the samples analyzed after six weeks (i.e., they were placed back in the freezer after analysis and analyzed again at twelve weeks). This was done to determine whether multiple thawing and freezing events affected nutrient concentrations. One-way ANOVAs were used to compare the nutrient concentrations between re-freezing events.

Results

Sampling Location

The average nutrient concentrations for the seven samples representing the cross section collected furthest downstream from the wastewater treatment plant (Below (II) WWTP) were first compared with five duplicate samples (Figures 5–9). The five duplicate samples were taken from a bucket collected from the middle of the stream also at the furthest location downstream from the wastewater treatment plant (Below (II) WWTP). Five filtered samples and five unfiltered samples were taken from the bucket. These samples were to represent samples that we would assume to be completely mixed. Because the five samples were collected from a mixed bucket, these samples were expected to be less variable. However, this was not observed for total nitrogen, nitrate, or soluble reactive phosphorus (Figures 6, 8, and 9). For these nutrients, the samples collected across the cross-section were less variable than the samples taken from the mixed bucket.

The Modified Bennet's test, the Wald test, and the Modified Miller test showed statistically significant differences between coefficients of variation for all nutrients between the four locations (Figure 1a), except for total nitrogen ($p > 0.05$) (Figure 10). However, when pairwise comparisons using nonparametric bootstrap analyses were computed for each nutrient (Figure 11), a statistical significance was observed for total nitrogen between two pairs of locations. Statistical significance was also observed between coefficients of variation for the four other nutrients between at least one pair of locations, these p-values are highlighted in red in Figure 11. Despite these differences, no pattern was observed across all nutrients. However, the coefficients of variation appeared to be lowest for samples collected at the point source in all nutrients compared to the other three locations. The nonparametric bootstrap analyses did not give all the expected results, seen by the blue value in Figure 11. A statistical significance was expected for this value because of the difference between the coefficients of variation between the two locations. An outlier in the data probably caused the observed results.

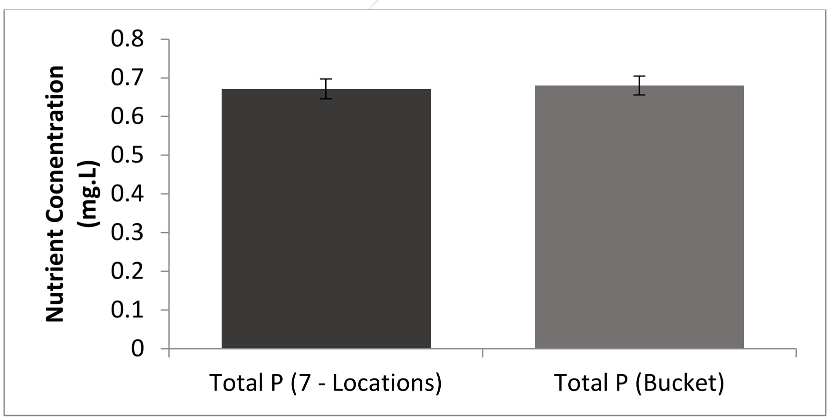


Figure 5. Comparison of total phosphorus concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

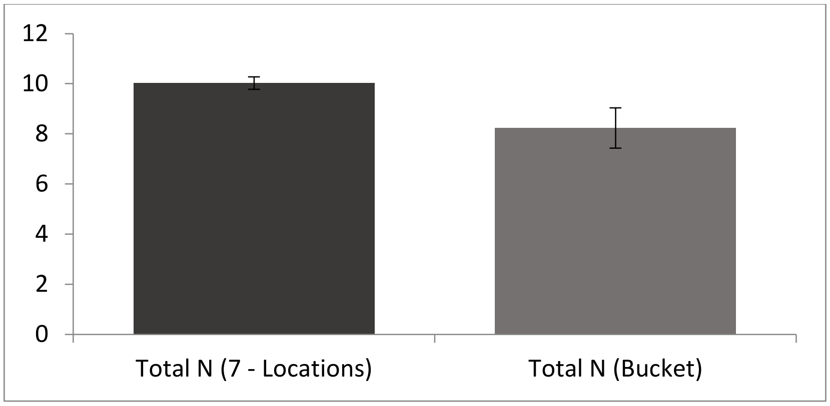


Figure 6. Comparison of total nitrogen concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

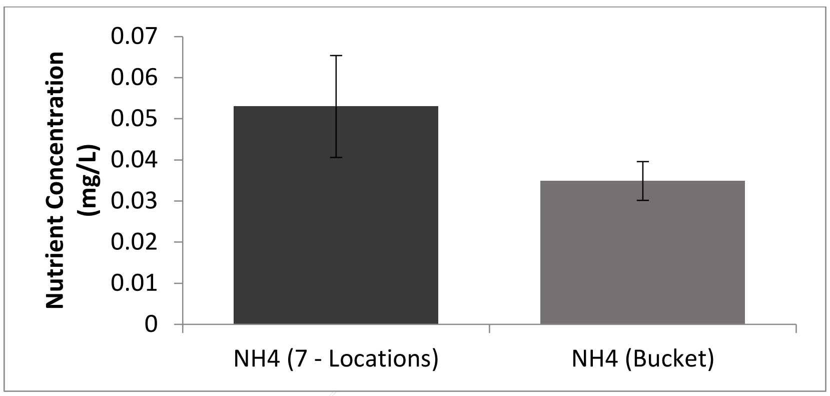


Figure 7. Comparison of ammonium concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

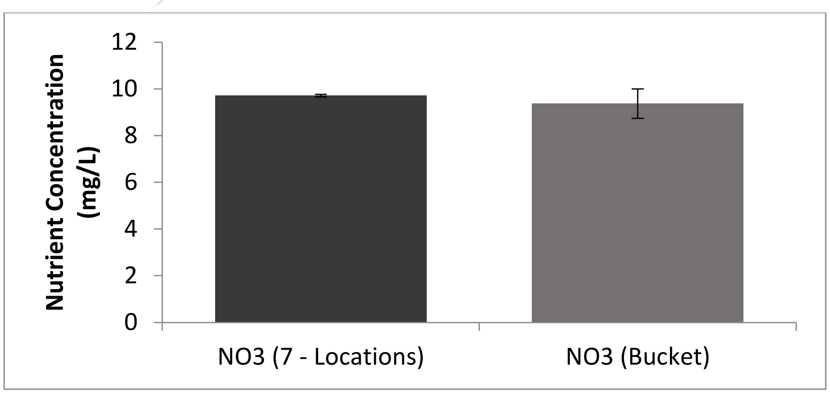


Figure 8. Comparison of nitrate concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

TP:

	P-Value
Modified Bennet's test	9.538705e-11
Wald test	7.770074e-07
Modified Miller test	2.340471e-12

TN:

	P-Value
Modified Bennet's test	0.06184338
Wald test	0.06317188
Modified Miller test	0.06858057

NH4:

	P-Value
Modified Bennet's test	0.02990244
Wald test	0.0453102
Modified Miller test	0.04590821

NO3:

	P-Value
Modified Bennet's test	9.94E-35
Wald test	4.75E-09
Modified Miller test	4.27E-29

SRP:

	P-Value
Modified Bennet's test	9.25E-21
Wald test	1.43E-08
Modified Miller test	1.88E-24

Figure 9. Comparison of soluble reactive phosphorus concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

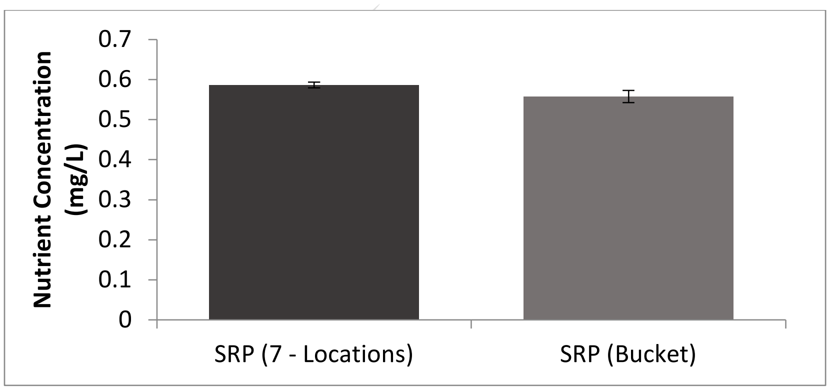


Figure 10. P-values computed from the Modified Bennet's test, the Wald test, and the Modified Miller test.

TP:

CV	0.2236685	0.02277014	0.08744049	0.1020543
	Above	Point Source	Below I	Below II
Above	x	0.02339532	0.1143771	0.1785643
Point Source	x	x	0.01939612	0.02579484
Below I	x	x	x	0.6540692
Below II	x	x	x	x

TN:

CV	0.06476794	0.1677015	0.1814677	0.06576692
	Above	Point Source	Below I	Below II
Above	x	0.08638272	0.0689862	0.7940412
Point Source	x	x	0.8408318	0.01779644
Below I	x	x	x	0.0269946
Below II	x	x	x	x

NH4:

CV	0.2517278	0.1308364	0.3704142	0.6181216
	Above	Point Source	Below I	Below II
Above	x	0.06778644	0.3537293	0.0679864
Point Source	x	x	0.09338132	0.2623475
Below I	x	x	x	0.4027195
Below II	x	x	x	x

NO3:

CV	0.6896826	0.01786509	0.02435554	0.01606915
	Above	Point Source	Below I	Below II
Above	x	0.01219756	0.01339732	0.0109978
Point Source	x	x	0.1269746	0.4727055
Below I	x	x	x	0.07218556
Below II	x	x	x	x

SRP:

CV	0.3125886	0.01331167	0.03949525	0.03206434
	Above	Point Source	Below I	Below II
Above	x	0.01539692	0.01639672	0.01259748
Point Source	x	x	0.02859428	0.0319936
Below I	x	x	x	0.2733453
Below II	x	x	x	x

Figure 11. P-values computed from nonparametric bootstrap analyses comparing two locations. Note that x is included in the table where no comparison is needed (i.e., the CV for above does not need to be compared to itself or where value is already provided in the table) and red values indicate a statistical significance.

Filtration Techniques

After conducting one-way ANOVAs using the average concentration for the twelve samples filtered with the geopump and the twelve samples filtered with a manual syringe (Figures 12–14), the only significant differences between nutrient concentrations were found in ammonium concentrations (Figure 12). The p-value determined for ammonium was 0.049, while the p-values for nitrate and SRP were 0.445 and 0.286, respectively. The ammonium concentrations were consistently higher in samples filtered with the syringe than with the pump. The difference between the average ammonium concentrations filtered with the syringe and with the pump was also greater than the method detection limit, verifying the statistical significance.

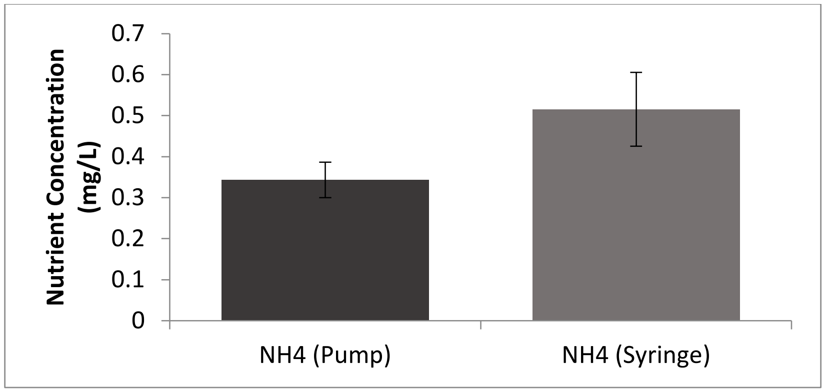


Figure 12. Comparison of ammonium concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

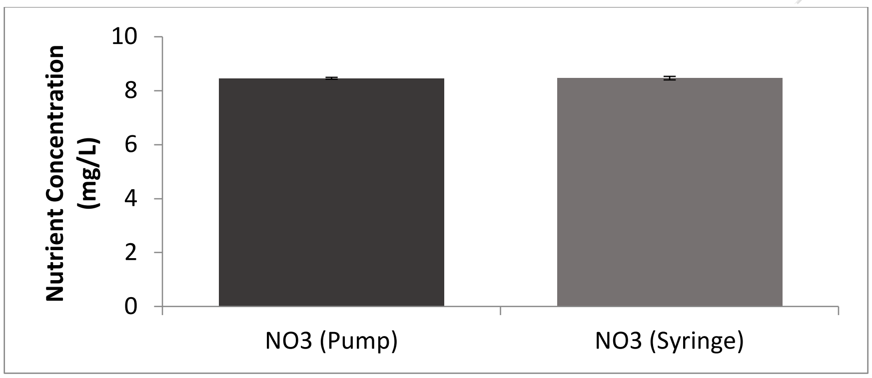


Figure 13. Comparison of nitrate concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

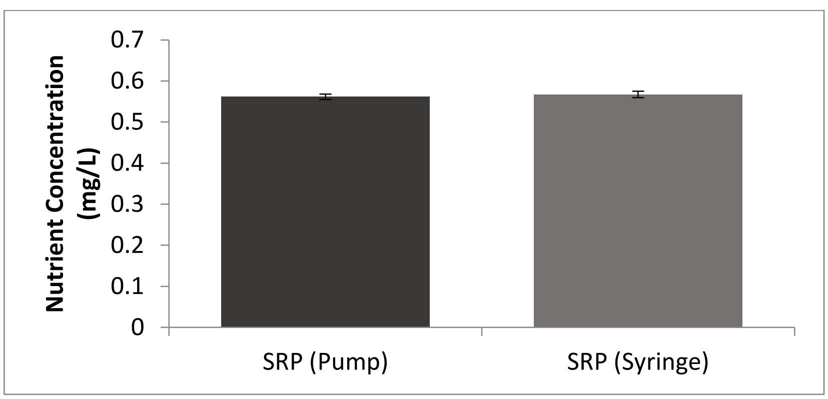


Figure 14. Comparison of soluble reactive phosphorus concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

Dilution Error

Comparisons of the dilutions I made and the dilutions made in the lab showed that the lab dilutions resulted in consistently lower nutrient concentrations for every nutrient (Figures 15–19). Standard errors for the dilutions I made were also consistently higher than standard errors for the lab dilutions. This may be due to less experience with making dilutions.

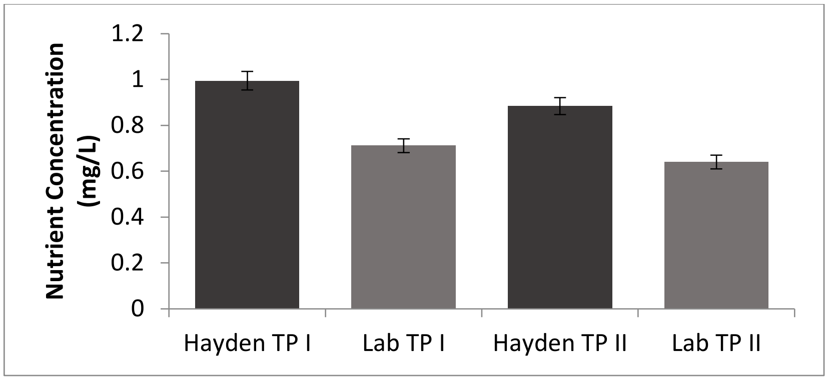


Figure 15. Comparison of total phosphorus concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

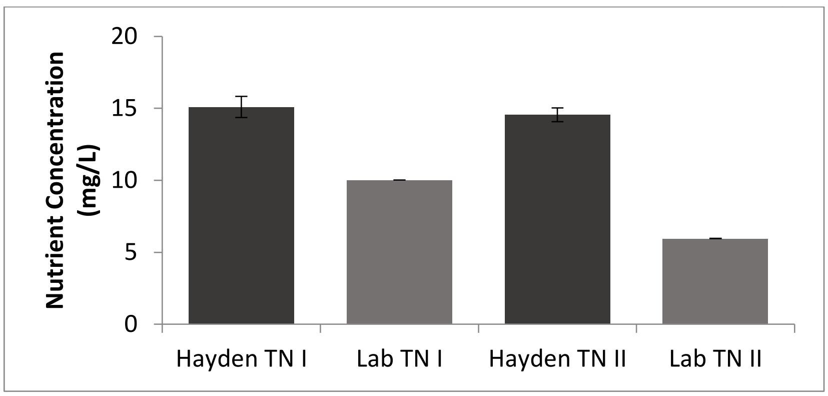


Figure 16. Comparison of total nitrogen concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

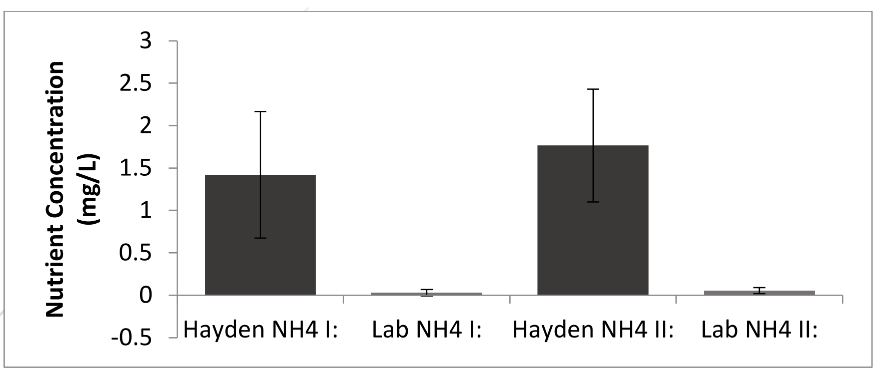


Figure 17. Comparison of ammonium concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

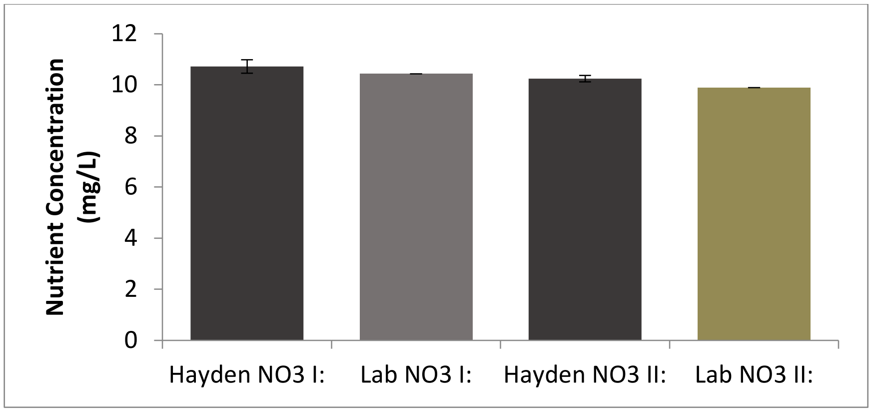


Figure 18. Comparison of nitrate concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

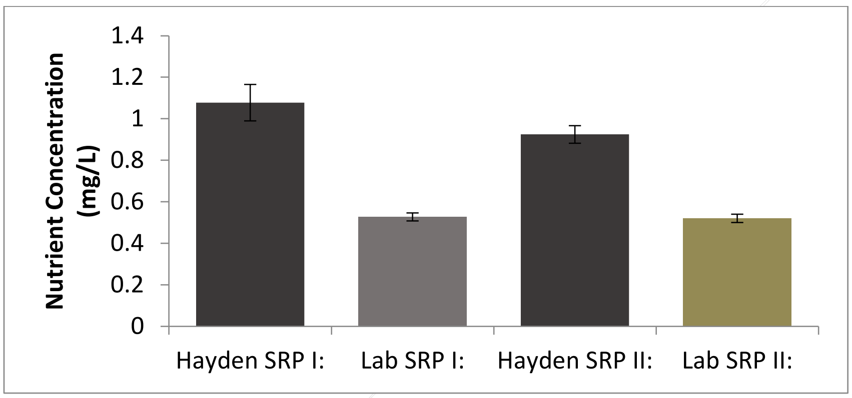


Figure 19. Comparison of soluble reactive phosphorus concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

Analytical Uncertainty

For each nutrient, the number of fails (QA/QC data beyond 20%) were tallied for both samples spiked and duplicate samples (Tables 1 and 2). Total nitrogen and ammonium incurred the greatest number of fails, but it most cases, the percent recovery and coefficients of variations were much less than 20%.

Table 3. Average coefficients of variation for sample duplicates and number of fails encountered for each nutrient.

	Average COV:	Number of Fails:
TN	4.2%	1
TP	4.5%	1
NH_4	-	4
NO_3	1.2%	-
SRP	6.5%	-

Table 4. Average percent recovery for spiked samples and number of fails encountered for each nutrient.

	Average % Recovery:	Number of Fails:
TN	97.9%	3
TP	101.3%	-
NH_4	93.5%	1
NO_3	93.4%	-
SRP	98.8%	-

Storage Time

Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations ($p < 0.05$) (Figures 20 and 21), however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates. The large concentration differences observed between total phosphorus concentrations of week one and the weeks following were, in large part, due to protocol failure (Figure 20). The samples analyzed on week one were locked in an autoclave overnight. The low concentrations on week one are attributed to this.

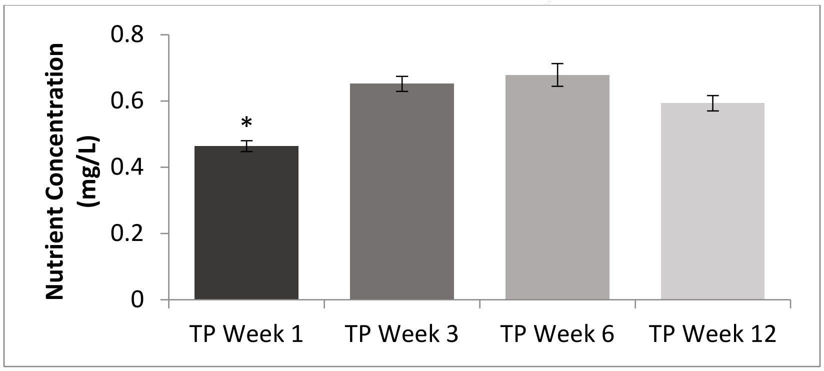


Figure 20. Average total phosphorus concentrations for samples analyzed at one week, three weeks, six weeks, and twelve weeks of freezing. The asterisk represents statistical significance due to lab protocol failure (i.e., samples being locked in an autoclave for loo long) and error bars represent standard error between samples.

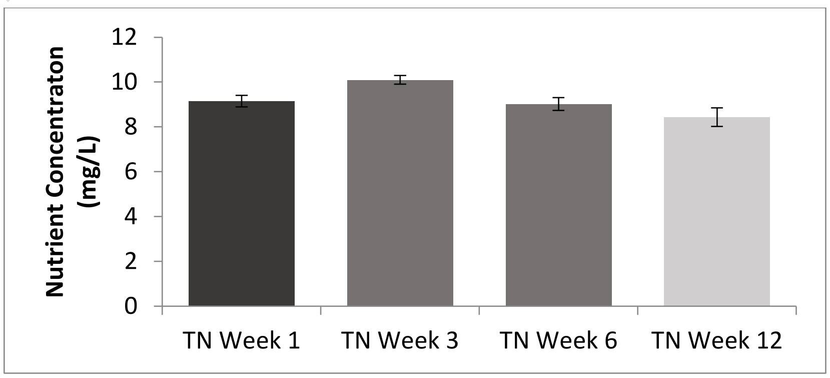


Figure 21. Average total nitrogen concentrations for samples analyzed at one week, three weeks, six weeks, and twelve weeks of freezing. The error bars represent standard error between samples.

When samples were thawed for initial analysis, put back in the freezer, and then thawed again for reanalysis, total nitrogen concentrations appeared to decrease, but total phosphorus seemed unaffected (Tables 3 and 4).

Table 5. Total phosphorus method detection limit and differences between samples thawed multiple times.

Total Phosphorus	mg/L
Method Detection Limit	0.43
Difference between samples opened once and opened twice	-0.007
Difference between samples opened twice and opened three times	0.03

Table 6. Total nitrogen method detection limit and differences between samples thawed multiple times.

Total Nitrogen	mg/L
Method Detection Limit	0.10
Difference between samples opened once and opened twice	1.59
Difference between samples opened twice and opened three times	0.91

Discussion

Nutrient mixing patterns between the four locations appeared to be different for each cross section. A general pattern was not observed; therefore mixing patterns are different for each nutrient between the four locations. According to Horowitz et al. (1990), poor selection of sampling locations within a cross-section could lead to inaccurate nutrient concentrations, due to over- or underestimation of these concentrations from variable mixing patterns. The coefficients of variation seemed to be lowest at the point source, and highest above the wastewater treatment plant, except for total nitrogen and ammonium concentrations. This could potentially lead to the hypothesis that higher concentrations confer less analytical variation; therefore concentration differences are masked by the overall high concentration.

When deciding whether to use an electric pump or a syringe when taking nutrient samples around a point source, caution should be used when analyzing for ammonium. From these data, it is not evident which filtering method is the more reliable one. When filtering with an electric pump, discretion should be used in order to assure that a tear or larger hole is not created in the filter because of too much pressure (Worsfold et al., 2005). Lambert et al. (1992) also observed the formation of filter cakes during filtration that led to changes in the effective pore size of the filter. Filters should be observed after use to determine if either of these events have occurred.

Dilutions made before analysis of nutrients were consistently lower than the dilutions I made, except for NO_2+NO_3-N. Variation was also greater in the dilutions I made, which is likely due to less experience. However, the differences observed between the dilutions I made and the dilutions made in the lab could be due to thawing samples multiple times before analysis. When making the dilutions, the samples were taken out of the freezer and thawed, and then placed back in the freezer. For analysis, the samples had to the thawed again. This is consistent with the data obtained from storage analysis, except total phosphorus concentrations also decreased, which was not observed in the storage analysis.

Analytical uncertainty proved to be less than uncertainty observed during sample collection and storage, except for unanticipated protocol failure. Protocol failure occurred due to samples being locked in an autoclave overnight before analysis, which caused some of the samples to completely evaporate. However, ammonium concentrations seemed to be the most variable.

Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations, however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates. These results are consistent with studies done by Avanzino and Kennedy (1993), Dore et al. (1996), and Venrick and Hayward (1985). However, according to Gordolinski et al. (2001), a standard storage protocol cannot be designed due to different chemical and biological characteristics of different sample matrices. This is one reason why some studies have found freezing to be an inadequate storage technique for some nutrients. The study done by Fellman et al. (2008) is an example of a study that determined that freezing was not an adequate storage technique for total dissolved phosphorus.

According to these data, appropriate sampling methods should be used when collecting samples within a cross-section. One option is using a composite sampling technique using an automatic water sampler to get a representative sample for the whole cross section (Facchi et al., 2007; Worsford et al., 2005). Martin et al. (1992) also observed different mixing patterns due to point source discharges, and recommended collecting grab samples at a “representative point” in a stream, if possible, or employing automatic water samplers. This may only be necessary in locations that have low nutrient concentrations, where the coefficients of variation seemed to be the greatest. Whitfield and McKinley (1981) observed that variability among field replicate samples was a great source of uncertainty in their study. This is consistent with the results of this study. Filtration methods should also be chosen appropriately when collecting samples for ammonium analysis. Finally, samples should not be thawed and frozen multiple times before analysis as this has been shown to decrease nutrient concentrations.

With these considerations in mind, the anomalies from the previous research were re-observed (Figures 22 and 23). Thawing samples multiple times could have accounted for a lower total nutrient concentration than constituent concentration if the multiple thawing only occurred in the samples analyzed for total concentrations, but not constituent concentrations. However, multiple thawing events could affect total nutrient concentrations differently than constituent concentrations, but this was not considered in this study. These anomalies could have also been due to a labeling error. The filtering technique is not a likely reason for the observed anomalies in the prior study because the anomalies are not consistent in time: samples where constituent nitrogen concentrations were higher than total nitrogen occurred early in the study, while constituent phosphorus samples were higher later in the sample collection. In addition, ammonium concentrations were low in comparison to nitrate+nitrite and total nitrogen concentrations and my research found that NH_4 could be influenced by filtration technique.

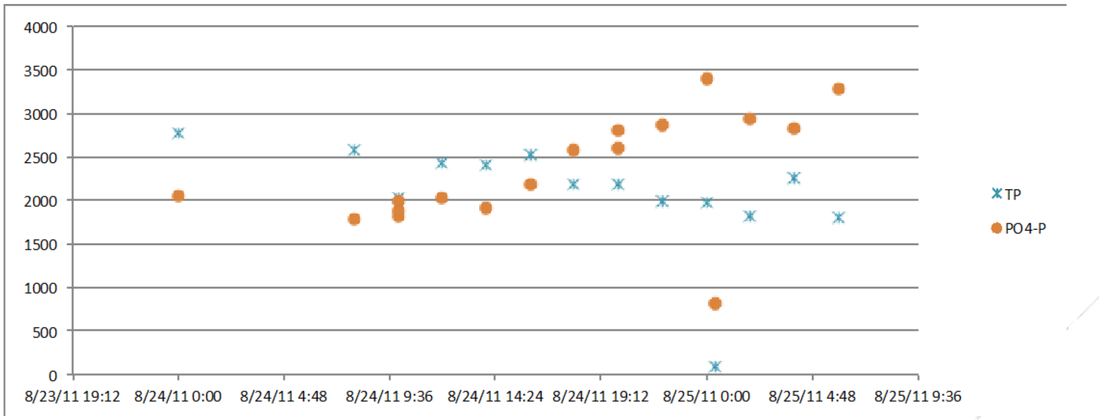


Figure 22. Anomalies observed in previous research. Phosphate concentrations were higher than total phosphorus concentrations in 7 of 13 samples.

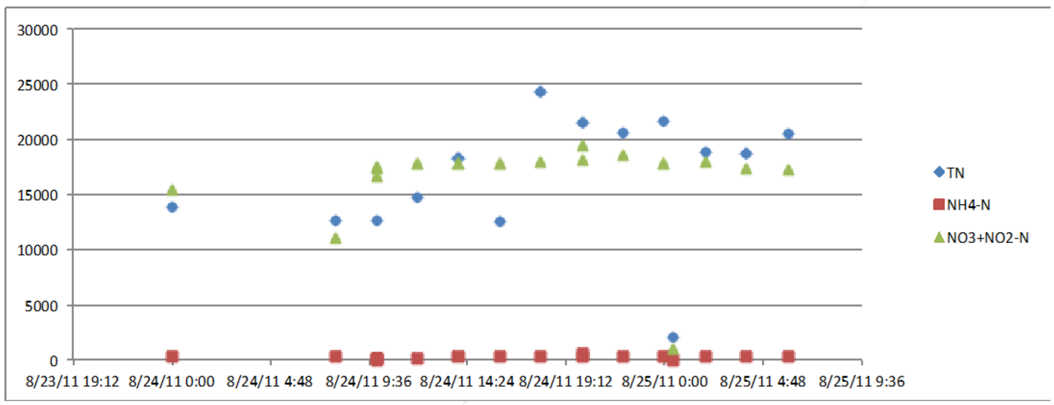


Figure 23. Anomalies observed in previous research. Nitrate+nitrite concentrations were higher than total nitrogen concentrations in 4 of 13 samples. However, the first sample may not be statistically different than the method detection level.

Future Research

Thawing samples multiple times proved to decrease total nitrogen concentrations, but not total phosphorus concentrations. This research did not test whether ammonium, nitrate, or soluble reactive phosphorus concentrations also decreased as a result of multiple thaws. More research in the future is needed to test the effects of multiple thaws on these constituent nutrient concentrations. A repeat sampling event could also be performed to determine if higher nutrient concentrations do confer less analytical variation as hypothesized above.

References

Avanzino, R. J., and V. C. Kennedy. 1993. Long-term frozen storage of stream water samples for dissolved orthophosphate, nitrate Plus nitrite, and ammonia analysis. Water Resources Research 29(10):3357–3362.

Dore, J. E., et al. 1996. Freezing as a method of sample preservation for the analysis of dissolved inorganic nutrients in seawater. Marine Chemistry 53:173–185.

Facchi, A., C. Gandolfi, and M. J. Whelan. 2007. A Comparison of River Water Quality Sampling Methodologies Under Highly Variable Load Conditions. Chemosphere 66(4):746–756.

Fellman, J. B., et al. 2008. An evaluation of freezing as a preservation technique for analyzing dissolved organic C, N and P in surface water samples. Science of the Total Environment 392(2–3):305–312.

Horowitz, A. J., et al. 1990. Variations in suspended sediment and associated trace element concentrations in selected riverine cross sections. Environmental Science & Technology 24(9):1313–1320.

Jafari, A. A., and M. R. Kazemi. 2013. A parametric bootstrap approach for the equality of coefficients of variation. Computational Statistics 28(6):2621–2639.

Lambert, D., et al. 1992. Changes in phosphorus fractions during storage of lake water. Water Resources 26(5):645–648.

Martin, G. R., et al. 1992. A comparison of surface-grab and cross sectionally integrated stream-water-quality sampling methods. Water Environment Research 64(7):866–876.

Venrick, E. L., and T. L. Hayward. 1985. Evaluation of some techniques for preserving nutrients in stored seawater samples. CalCOF Report, 26,160–168.

Whitfield, P. H., and J. W. McKinley. 1981. Some factors affecting the determination of particulate carbon and nitrogen in river water. Journal of the American Water Resources Association 17(3):381–386.

Worsfold, P., et al. 2005. Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils. Talanta 66(2):273–293.