Final Report to Governors from the Joint Study Committee and Scientific Professionals

Summary and Recommendations

The committee met between October 2013 and December 2016, selected qualified scientific professionals, developed a scope of work, completed the 2-year joint study, reviewed the results and used a weight of evidence approach to recommend a six-month average total phosphorus level of not to exceed 0.035 milligrams per liter based on water samples collected during critical conditions was necessary to protect the designated [Oklahoma] Scenic Rivers.

Respectfully Approved December 19, 2016

Brian Haggard, PhD
Co-Chair
Arkansas

Derek Smithee
Co-Chair
Oklahoma

Ryan Benefield
Arkansas

Shellie Chard
Oklahoma

Marty Matlock, PhD
Arkansas

Shanon Phillips
Oklahoma

Ryan King, PhD
Baylor University

Technical Summary and Recommendations

Introduction

The six Oklahoma Scenic Rivers, particularly the Illinois River Watershed, have been a focus of conservation and management efforts to improve water quality by Arkansas and Oklahoma. The Illinois River Watershed is a trans-boundary watershed in the Ozark Plateaus with its headwaters in northwest Arkansas, and this watershed includes three of the designated Oklahoma Scenic Rivers – the Illinois River, Flint Creek and Baron Fork. The other Oklahoma Scenic Rivers include Little Lee Creek and Lee Creek in the watershed to the south of the Illinois River Watershed, as well as the Mountain Fork further south in the Ouachita Mountains. However, the focus of the environmental issues, elevated phosphorus (P) concentrations in the streams and rivers, and management have centered on the trans-boundary Illinois River Watershed.

In 2003, the states signed the [first] ***Joint Statement of Principles and Actions*** stating the shared goal of improving water quality in the Illinois River Watershed, resulting in effluent total phosphorus (TP) limits of 1 mg L^{-1} on municipal facilities with a design capacity of greater than 1 million gallons per day (MGD) and Arkansas passing legislation and regulations on poultry litter management. The management changes in the Illinois River Watershed improved water quality, reducing phosphorus concentrations and loads in the Illinois River (Haggard, 2010; Scott et al., 2011). The changes in TP concentrations and loads were subsequent to changes in effluent P inputs from one facility upstream in northwest Arkansas [i.e., Springdale's wastewater treatment plan (WWTP)] (Scott et al., 2011) to which elevated TP concentrations could be traced upstream (see Haggard, 2010).

However, TP concentrations in the Illinois River and select tributaries were still greater than the numeric TP criteria (0.037 mg L^{-1}, OWRB, 2002, OAC 785:45) applicable to Oklahoma's Scenic Rivers seasonally in 2009 and dependent upon flow conditions (Scott et al., 2011). Continuing in a collaborative fashion, the states then adopted a ***Second Statement of Joint Principles and Actions*** (hereafter, ***Second Statement***) in 2013 augmenting the first agreement, providing a three-year extension of commitments. The premise of the Second Statement included the governors' appointment of six individuals to the "**JOINT STUDY COMMITTEE**" who were required to reach agreement on the procurement, execution and conduct of the "**JOINT STUDY**" as defined with the terms of the *Second Statement*. The costs (i.e, $600,000) of the **JOINT STUDY** were paid for by Arkansas parties and funds placed in repository with the Arkansas-Oklahoma Arkansas River Compact Commission. The **JOINT STUDY COMMITTEE** was authorized to formulate the scope of work and select qualified scientific professionals (who do not reside in nor principal business locations within the states) to conduct the **JOINT STUDY**.

The **JOINT STUDY** included mandatory components as defined in the *Second Statement* which guided the formation of the scope of work by the **JOINT STUDY COMMITTEE** and selected contractor, i.e. qualified scientific professionals. The three important mandatory components included:

(1) "The primary purpose of the **JOINT STUDY** is to determine the TP threshold response level, in mg L^{-1}, at which any statistical shift occurs in algal species composition or algal biomass

production resulting in undesirable aesthetic or water quality conditions in the Designated Scenic Rivers."

(2) "The **JOINT STUDY** shall be completed in accordance with U.S. EPA Rapid Bio-assessment Protocols… and follow EPA's most recent guidance 'Using Stressor-response Relationships to Derive Numeric Nutrient Criteria…" (EPA, 2010).

(3) "The **JOINT STUDY** shall include a sampling population that is adequate to determine the frequency and duration component of the numeric criterion."

The **JOINT STUDY COMMITTEE** issued a Request for Qualifications (RFQ), interviewed three professional teams, and then selected Dr. Ryan King's research group at Baylor University to perform the negotiated scope of work specific to the **JOINT STUDY**. All Statement of Qualifications (SOQs), meeting minutes, interim reports and reference materials are available on the web at:

www.ok.gov/conservation/Agency_Divisions/Water_Quality_Division/IR_Joint_Study_Committee.html

The purpose of this report is to provide "an objective analysis of the water quality data" and identifies the relation between TP concentrations and "multiple ecological response levels" targeted at protecting the Oklahoma Scenic Rivers from "undesirable aesthetic and water quality conditions." The **JOINT COMMITTEE** unanimously made "specific recommendations as to what TP levels, and what frequency and duration components of measure, are necessary to protect the aesthetics beneficial use and scenic river (Outstanding Water Resource) designations assigned to the designated [Oklahoma] Scenic Rivers" based on the relation between TP concentrations and "biotic indicators of water quality, including primarily algal taxonomic composition and periphyton biomass." The technical report from the selected scientific professionals is provided as an appendix of to this report, and it provides the expansive details of the sampling, data collected, statistical analysis, and additional supplemental information.

Joint Study Methods and Data Analysis

The sampling sites selected for the **JOINT STUDY** targeted "streams and rivers within the same EPA eco-region and comparable to the streams in the designated Scenic River watershed in terms of stream order and watershed land uses." A total of 35 stream reaches were selected for the **JOINT STUDY**, and the majority of the stream reaches were within five of the six designated Scenic River watersheds, including the Illinois River, Flint Creek, Baron Fork, Little Lee Creek and Lee Creek watersheds. Additionally, stream reaches were also included in adjacent watershed within the same EPA eco-region. The stream reaches were selected based on these criteria: (1) presence of riffles, (2) cobble substrate (10-20 cm), (3) open tree canopy, and (4) fast, turbulent flow. The ultimate goal of the site selection was to have stream reaches or sites with a gradient of TP concentrations sufficient to evaluate thresholds in algal taxa and biomass response with increasing TP concentrations.

Water and biological sampling occurred on an every other month schedule, subject to flow conditions, from June 2014 through April 2016. Water samples were collected at the upstream boundary of each stream reach and then analyzed for TP and other water-quality parameters at the Baylor University Center for Reservoir and Aquatic Ecosystems Research (CRASR) following standard methods and approved quality assurance and quality control protocols. Periphyton was removed from 15 cobbles in the desired size class (10-20 cm) from each stream reach and analyzed for periphyton biomass [mg chlorophyll-a (chl-a) m^{-2}] and algal species composition of diatoms and soft algae. The diatoms and soft algae were enumerated by species and reported as biovolume. Sampling was successfully completed every other month during base flow conditions at all 35 stream reaches or sites over the two-year study, with the exception of two sites where the stream was not flowing in October 2014 and one site observed in backwater conditions (i.e., flooded by Lake Tenkiller) during June 2015 and December 2015. Water and biological samples were collected over a variety of flows across the study (see Appendix, Figure 9), including relatively low conditions and following historic flooding in late December 2015. The **JOINT STUDY COMMITTEE** unanimously defined the '**CRITICAL CONDITIONS**' for the **JOINT STUDY** as the conditions where surface runoff is not the dominant influence of total flow and stream ecosystem processes.

The **JOINT STUDY** used various statistical techniques to analyze for TP thresholds with algal species composition (i.e., biovolume) and periphyton biomass (mg chl-a m^{-2}). The main techniques employed included:

(1) A nonparametric form of change point analysis (nCPA, King and Richardson, 2003) was used to determine threshold in periphyton biomass and select algal species (i.e., Cladophora biovolume). This statistical technique estimates the probability that the variance in the data explained by the model (i.e., threshold) is not better than expected by chance and provides estimates of uncertainty (i.e., confidence intervals) about where the true threshold might be. This technique is recommended for deriving numeric nutrient criteria (see EPA, 2010).

(2) Threshold Indicator Taxa Analysis (TITAN, Baker and King, 2010), which is an analytical approach used to identify thresholds among many algal species simultaneously in in response to a stressor gradient (i.e., increasing TP concentrations; the details of this technique are available in the appendix. TITAN provides TP threshold information on individual species, as well as community-level responses, that is, the groups of algal organisms that are decreasing (Sumz-) in abundance (i.e., biovolume) and increasing (Sumz+) in abundance across the TP concentration gradient. TITAN also provides uncertainty or confidence intervals about the TP threshold.

The final technical report (see Appendix) further outlines additional statistical techniques that were used in the **JOINT STUDY**, providing additional weight of evidence to support the recommendation put forth by the **JOINT STUDY COMMITTEE**. The **JOINT STUDY COMMITTEE** unanimously agreed that the range in TP thresholds from the **JOINT STUDY** were developed based on water and biological samples collected under **CRITICAL CONDITIONS**.

Joint Study Results

Phosphorus Thresholds with Periphyton Biomass and Nuisance Algal Taxa

TP concentrations were relatively consistent within each site over time, although several sites showed seasonal variability associated with dilution of effluent inputs. The concentrations were also depressed below the median TP concentrations over the 2-year **JOINT STUDY** during select samplings when algal biomass and primary production were high. Overall, TP concentrations in individual water samples ranged from less than 0.01 mg L^{-1} to almost 0.20 mg L^{-1} (see Appendix, Figure 10), and 2-year study averages varied from less than 0.01 mg L^{-1} to greater than 0.10 mg L^{-1} (at two sites or stream reaches). The evidence presented in the **JOINT STUDY** showed that a focus on TP as the potential driver of potential nuisance conditions of biomass and algal species composition was supported.

Benthic chl-a varied over time among the study sites or stream reaches, as well as within an individual site when higher productivity often existed. The average benthic chl-a across the 35 sites or stream reaches over the 2-year **JOINT STUDY** varied from ~50 mg m^{-2} to over 600 mg m^{-2}, while the benthic chl-a measured at discrete samplings varied from less than 50 mg m^{-2} across several sites to over 1000 mg m^{-2} (see Appendix, Figures 13-14). The dramatic increases in benthic chl-a observed in two sampling months (i.e., December 2014 and February 2015) coincided with blooms of *Cladophora glomerata* (hereafter, *Cladophora*).

The scientific professionals analyzed the relations between benthic chl-a and TP concentrations over a variety of durations (from 2 to 12 months), producing over 110 TP thresholds for the **JOINT STUDY COMMITTEE** to evaluate. The **JOINT STUDY COMMITTEE** agreed to put more weight on average TP concentrations over a 6 month duration or longer period, providing still almost 70 different TP thresholds with periphyton biomass for consideration. The TP concentrations were a statistical significant shift in periphyton biomass occurred varied from 0.014 to 0.060 mg L^{-1} with instantaneous benthic chl-a and from 0.018 to 0.040 mg L^{-1} for average benthic chl-a over the same duration (see Appendix, Tables 4-5, Figure 16-17). The average benthic chl-a across all sites above the TP threshold (i.e., sites with TP concentrations greater than the change point) was 2 or more times greater than average benthic chl-a at all sites with TP concentrations less than the threshold.

The dominant filamentous algae was *Cladophora*, which is widely known as a nuisance species that increases in abundance [essentially biovolume] with nutrient enrichment (Dodds and Gudder, 1992). *Cladophora* was not present to very low in biovolume at sites or stream reaches with relatively low TP concentrations, but showed a non-linear change in biovolume as TP concentrations increased across the sampling locations. The scientific professionals recommended that the **JOINT STUDY COMMITTEE** focus on mean responses of *Cladopora* biovolume to the increasing TP gradient, because of measurement variability with soft, filamentous algae. The TP thresholds showing an increase in average *Cladophora* biovolume across all sites and at least a six-month duration varied from 0.032 to 0.051 mg L^{-1}, with 16 out of 17 change points evaluated being 0.035 mg L^{-1} or greater (see Appendix, Tables 6, Figure 18).

The **JOINT STUDY COMMITTEE** and scientific professionals also evaluated how the proportion of total biovolume of nuisance algal taxa changed across the increasing TP gradient, where five genera of filamentous green algae that occurred in our data set were classified as nuisance taxa: *Cladophora, Oedogonium, Rhizoclonium, Spirogyra, and Hydrodictyon.* However, *Cladophora* was the dominant species of the total nuisance biovolume (generally greater than 95%); there were a few sites that had blooms of other algal species across the 2-year **JOINT STUDY**. The **JOINT STUDY COMMITTEE** per scientific professional recommendation focused on average proportions of the nuisance algal taxa over durations of six months or longer. The analysis showed that significant TP thresholds were present in 15 out of 17 relations evaluated, where TP thresholds ranged from 0.033 to 0.058 mg L^{-1} with 14 out of 15 TP thresholds being at concentrations of 0.035 mg L^{-1} or greater (see Appendix, Table 7, Figure 20).

Phosphorus Thresholds with TITAN Analysis

The community level analysis of algal species response to an increasing TP gradient provided additional information, and it was considered in the weight of evidence used to make recommendations (see Appendix, Tables 8-9, Figure 21). When looking at community level responses, the **JOINT STUDY COMMITTEE** evaluated the change in average taxa biovolumes over six month durations or longer in TITAN. Various algal species declined in abundance (measured as biovolume) as TP concentrations increased, where mean community level shifts in the natural assemblage of algae occurred at TP concentrations as low as 0.011 mg L^{-1} to as high as 0.049 mg L^{-1}. The algal species that declined (Sumz-) in abundance had a lower range in TP thresholds, where mean cumulative shifts occurred at TP concentrations from 0.011 to 0.025 mg L^{-1}. On the other hand, the algal species that increased (Sumz+) in abundance had TP thresholds that over lapped with the Sumz- scores, ranging from TP concentrations of 0.019 to 0.049 mg L^{-1}. TITAN analysis also shows change points (i.e., TP thresholds) for individual species, and the TP thresholds based on TITAN for *Cladorphora* were within the range reported above specific to nCPA analysis (0.032–0.051 mg L^{-1}) but were on the lower end of this range.

Joint Study Committee Recommendations

The **JOINT STUDY COMMITTEE** met ten times between October 2013 and December 2016, where all meeting were open to the public and information including agendas, minutes, and interim reports were posted on the web site dedicated to this committee; there were eight interim reports prepared and presented by the employed scientific professionals, i.e. Dr. Ryan King, Baylor University. The **JOINT STUDY COMMITTEE** unanimously agreed that the **JOINT STUDY** was performed and provided data to meet the '**CHARGE**' of the **JOINT STUDY COMMITTEE** as defined in the third paragraph of page 3 under 'USE OF STUDY FINDINGS AND RESULTS.' The **CHARGE** was "…to make specific recommendations as to what TP levels, and what frequency and duration components of measure, are necessary to protect the aesthetics beneficial use and scenic river (Outstanding Resource Water) designations assigned to the designated Scenic Rivers." The **CHARGE** goes on to state that the recommendation of the **JOINT STUDY COMMITTEE** will be "…based on overall stream health which shall include evaluating the relationship, if

any' between TP concentrations… and biotic indicators of water quality, including primarily algal taxonomic composition and periphyton biomass."

The **JOINT STUDY COMMITTEE** unanimously agreed on several key, factual elements based on the TP thresholds identified in the **JOINT STUDY** and briefly discussed above, including:

(1) The **JOINT STUDY** showed the change in algal taxonomic composition and periphyton biomass was statistically observed at TP concentrations as low as 0.011 mg L^{-1} and as high as 0.074 mg L^{-1}. [Note: This was based on all thresholds reported in the appendix.]

(2) The **JOINT STUDY** showed that statistical shifts in mean *Cladophora* biovolume and mean nuisance taxa proportion of total biovolume was observed between 0.032 and 0.058 mg TP L^{-1}.

(3) The **JOINT STUDY** showed that the largest mean cumulative shift in the natural assemblage of algal species was observed within the range from 0.011 to 0.049 mg TP L^{-1} where species declined in abundance within the range from 0.011 to 0.025 mg TP L^{-1} and species increased in abundance within the range from 0.019 to 0.049 mg TP L^{-1}.

The **JOINT STUDY COMMITTEE** considered the plethora of scientific evidence and statistical analysis provided by the **JOINT STUDY**, but the focus was on the TP concentration thresholds with regard to nuisance algal species (i.e, Cladophora biovolume and nuisance taxa proportion). The **JOINT STUDY COMMITTEE** and its scientific professionals (Dr. Ryan King) employed to complete the **JOINT STUDY** specifically and unanimously recommend:

> *A six-month average total phosphorus level of not to exceed 0.035 mg L^{-1} based on water samples taken during the CRITICAL CONDITION, as previously defined, was necessary to protect the aesthetics beneficial use and scenic river (Outstanding Resource Water) designations assigned to the designated Scenic Rivers.*

The **JOINT STUDY COMMITTEE** also discussed at length how the recommended TP threshold (0.035 mg L^{-1} under defined conditions) related to the periphyton biomass based on generalized additive modeling (GAM, see Appendix, Tables 10, Figures 23-24), and then how predicted periphyton biomass compared to benthic chl-a thresholds where *Cladorphora* biovolume increased significantly (see Appendix, Figure 22). However, the **JOINT STUDY COMMITTEE** put more weight on the TP thresholds associated with *Cladophora* biovolume and proportion of nuisance algal taxa (relative to total biovolume) in discussion specific to making a recommendation to meet the **CHARGE** of the ***Second Statement***. The **JOINT STUDY** provided "reliable and objective data analysis that will then form the basis for the Parties and EPA to make informed decisions about the scientific merit of any proposed revisions to the TP criterion for the designated Scenic Rivers."

Furthermore, the **JOINT STUDY COMMITTEE** unanimously recommends that the states (Arkansas and Oklahoma) develop a monitoring and assessment program informed by the **JOINT STUDY** and other scientific information to determine attainment of the criteria.

Finally, the **JOINT STUDY COMMITTEE** unanimously recommends that protection of the [Oklahoma] Scenic Rivers needs to extend beyond the phosphorus levels and additionally focus on including but limited to the following:

- Hydrologic alteration
- Riparian zone protection
- Stream bank stabilization
- Fluvial channel habitat
- In-stream mining
- And, other contaminants.

And, the **JOINT STUDY COMMITTEE** unanimously views system wide management as critical to the protection of the [Oklahoma] Scenic Rivers.

References

Dodds, W.K., and D.A. Gudder. 1992. The ecology of *Cladophora*. Journal of Phycology 28:415–427.

EPA. 2010. Using Stressor-response Relationships to Derive Numeric Nutrient Criteria.

Haggard, B.E. 2010. Phosphorus concentrations, loads and sources within the Illinois River drainage area, northwest Arkansas, 1997–2008. Journal of Environmental Quality 39:2113–2120.

Scott, J.T., B.E. Haggard, A.N. Sharpley, and J.J. Romeis. 2011. Change point analysis of phosphorus trends in the Illinois River (Oklahoma) demonstrates the effects of watershed management. Journal of Environmental Quality 40:1249–1256.

Appendix

The next pages contain the final report submitted by Dr. Ryan King to the **JOINT STUDY COMMITTEE** to fulfil his obligations in the completion of the **JOINT STUDY** and the contract with Baylor University.

Oklahoma-Arkansas Scenic Rivers Joint Phosphorus Study

FINAL REPORT

19 December 2016

Principal Investigator:
Ryan S. King
Professor, Department of Biology
Baylor University
Waco, TX 76798
www.baylor.edu/aquaticlab

Joint Study Committee Members:
Brian Haggard; Co-Chair (University of Arkansas)
Marty Matlock (University of Arkansas)
Ryan Benefield (Arkansas Natural Resources Commission)
Derek Smithee; Co-Chair (Oklahoma Water Resources Board)
Shellie Chard (Oklahoma Dept. of Environmental Quality)
Shanon Philips (Oklahoma Conservation Commission)

Study Framework

The Oklahoma-Arkansas Scenic Rivers Joint Phosphorus study was executed in accordance with the *Second Statement of Joint Principles and Actions*. The primary purpose of this study was (p.2, *Mandatory Study Components*):

"to determine the total phosphorus threshold response level....at which any statistically significant shift occurs in

1. algal species composition or
2. algal biomass production

...resulting in undesirable

1. aesthetic or
2. water quality

...conditions in the Designated Scenic Rivers."

Furthermore (p.3-4, *Use of Study Findings and Results*):

"The States of Arkansas and Oklahoma, acting through their respective Parties, agree to be bound by the findings of the Joint Study. Oklahoma, through the Oklahoma Water Resources Board, agrees to promulgate any new Numeric Phosphorus Criterion, subject to applicable Oklahoma statutes, rules and regulations if significantly different than the current 0.037 mg/L standard. "Significantly different" means the new Numeric Phosphorus Criterion exceeds -.010 or +.010 than the current .037 criterion. If the new Numeric Phosphorus Criterion is at or between .027 and .047, then the State of Oklahoma is not required to promulgate the new criterion in its water quality standards. Arkansas agrees to be bound by and to fully comply with the Numeric Phosphorus Criterion at the Arkansas-Oklahoma State line, whether the existing 0.037 mg/L standard is confirmed or a new Numeric Phosphorus Criterion is promulgated. Parties for the States of Arkansas and Oklahoma shall forego any legal or administrative challenges to the Joint Study."

This report summarizes the work performed by Baylor University (the third party contractor) along with Joint Study Committee in the context of this study framework. The results presented herein are based on a field gradient "stressor-response" study designed to identify levels of total phosphorus that lead to the undesired outcomes described above. The study design, site selection, measurement endpoints, field methods, and statistical analyses were vetted and unanimously approved by the 6-member Joint Study Committee. Further, the results presented correspond to specifically requested analyses by members of the Joint Study Committee. This report does not include recommendations or conclusions regarding the numerical criterion. This report serves to guide the Joint Study Committee towards an informed, scientifically grounded recommendation for a numerical phosphorus criterion for the Oklahoma Scenic Rivers based on the results herein.

Study Design

Site selection

Thirty-five stream reaches were selected for the study. These sites were located in watersheds of 5 of the 6 Oklahoma Designated Scenic Rivers (Illinois River, Flint Creek, Barren Fork Creek, Little Lee Creek, and Lee Creek; Table 1, Figure 1). The Joint Study committee elected to exclude the Mountain Fork River for logistical reasons.

Candidate reaches were selected based on the following characteristics: (1) presence of riffle channel unit(s); (2) predominance of medium-to-large cobble substrate (10-20 cm); (3) mostly to fully open tree canopy (full sun), and (4) fast, turbulent flow, which is not always a characteristic of riffles in small streams but is in larger streams and rivers that were the primary focus of this study. The combination of these factors was deemed critical to ensure comparability between smaller streams and rivers in the study region and the Illinois River, the largest river in the study. The mainstem Illinois typically had habitat that met all four of these criteria, thus reaches included in the study from other rivers and streams had to also meet these criteria. For example, had we sampled a subset of streams that had only gravel substrate in their riffles, the results would have been confounded by the fact that gravel is scoured much more easily than cobble because even the slightest changes in flow cause these substrates to roll downstream. Nuisance filamentous algae such as *Cladophora* are much more likely to be collected on larger, more stable substrates, and, when coupled with turbulent flow, are the typical locations where nuisance algal blooms are initiated in the large streams and rivers (Dodds and Gudder 1992). Canopy cover also was important because all of the Illinois River mainstem sites were open canopy and very low light conditions associated with dense tree canopy would have limited algal growth and confounded comparisons to open-canopy sites on the Illinois and other large streams in the study area.

Reaches that met these criteria were prioritized for selection if they (1) had an existing USGS stream gage at or near the site, (2) had been or were being monitored for nutrients by Oklahoma or Arkansas. Additionally, the committee prioritized sites on the Illinois River because of its high levels of recreational use and socioeconomic importance to the region.

Reaches were excluded if obvious gravel extraction activity, construction, or anything unusual at or near the site that could have affected the potential relationship between phosphorus and biological response variables were evident.

If all of these conditions were met, the final, most important criterion for site selection was that the sites spanned a gradient of total phosphorus (TP) representative of the full range of TP conditions in the Scenic Rivers, their tributaries, and adjacent watersheds. Existing TP data from intensively monitored locations by the University of Arkansas, Oklahoma Water Resources Board, and Oklahoma Conservation Commission guided the initial screening of sites for inclusion in the gradient study, along with an extensive sampling of 60 sites in April 2014 to identify additional locations not previously studied by these organizations. Based on these data, 35 stream reaches were chosen. Each site filled a gap in the continuum of total phosphorus

concentrations from the lowest to the highest in the region such that the distribution of TP among sites was roughly log-linear.

Table 1. Site codes, coordinates, and location description of the 35 stream reaches.

Name	Latitude	Longitude	Description
BALL1	36.06137	-94.5732	Ballard @ E0660 Rd
BARR1	35.87954	-94.4822	Barren Fk @ SH45 Dutch Mills
BARR2	35.91906	-94.6193	Barren Fk @ SH59 nr Baron
BARR3	35.94727	-94.6935	Barren Fk @ N4670 Rd Christie
BARR4	35.87013	-94.897	Barren Fk @ Welling Br
BEAT1	36.35495	-94.7767	Beaty @ D0458 Rd
CANE1	35.78497	-94.8559	Caney @ Welling Road
COVE1	35.68576	-94.3663	Cove @ Creek Fk Rd
EVAN1	35.87742	-94.5706	Evansville @ D0795 Rd.
FLIN1	36.23973	-94.5007	Flint @ Dawn Hill East Rd nr. Gentry
FLIN2	36.21771	-94.6019	Flint @ D0553 nr West Siloam Springs
FLIN3	36.21454	-94.6655	Flint @ D4680 Rd Hazelnut Hollow
GOOS1	36.05603	-94.2912	Goose @ Little Elm Rd CR19
ILLI1	35.95398	-94.2494	Illinois @ Orr Rd
ILLI2	36.10135	-94.3441	Illinois @ SH16 nr Savoy
ILLI3	36.16864	-94.4355	Illinois @ Chambers Springs Rd
ILLI4	36.1093	-94.5339	Illinois @ SH 59 AR Canoeing
ILLI5	36.14201	-94.6681	Illinois @ N4695 low water xing & River Rd
ILLI6	36.17349	-94.7237	Illinois @ Flint Cr
ILLI7	36.06755	-94.8823	Illinois @ Hanging Rock SH10
ILLI8	35.91667	-94.928	Illinois @ SH62 Tahlequah
LEE1	35.68091	-94.3578	Lee @ Creek Fk Rd
LLEE1	35.57263	-94.5567	Little Lee @ SH101 Nicut
LSAL1	36.28455	-95.0887	Little Saline @ E506 Rd
MTFK1	35.68016	-94.4558	Mountain Fk @ SH59 pulloff S of Davidson
OSAG1	36.26593	-94.2378	Osage @ Healing Springs Rd CR264
OSAG2	36.222	-94.2901	Osage @ Snavely Rd
SAGE1	36.198	-94.5829	Sager @ Beaver Springs Rd.
SALI1	36.28154	-95.0932	Saline @ E6508 Rd USGS site
SPAR1	36.24367	-94.2393	Spring @ SH112 AR
SPAV1	36.38485	-94.481	Spavinaw @ Limeklin Rd CR29
SPAV2	36.32323	-94.6854	Spavinaw @ Colcord Kiethy Rd
SPRG1	36.1429	-94.9091	Spring @ Rocky Ford Rd & N556
SPRG2	36.09092	-95.0147	Spring @ N485 Rd low water xing
SPRG3	36.14833	-95.1548	Spring @ SH82

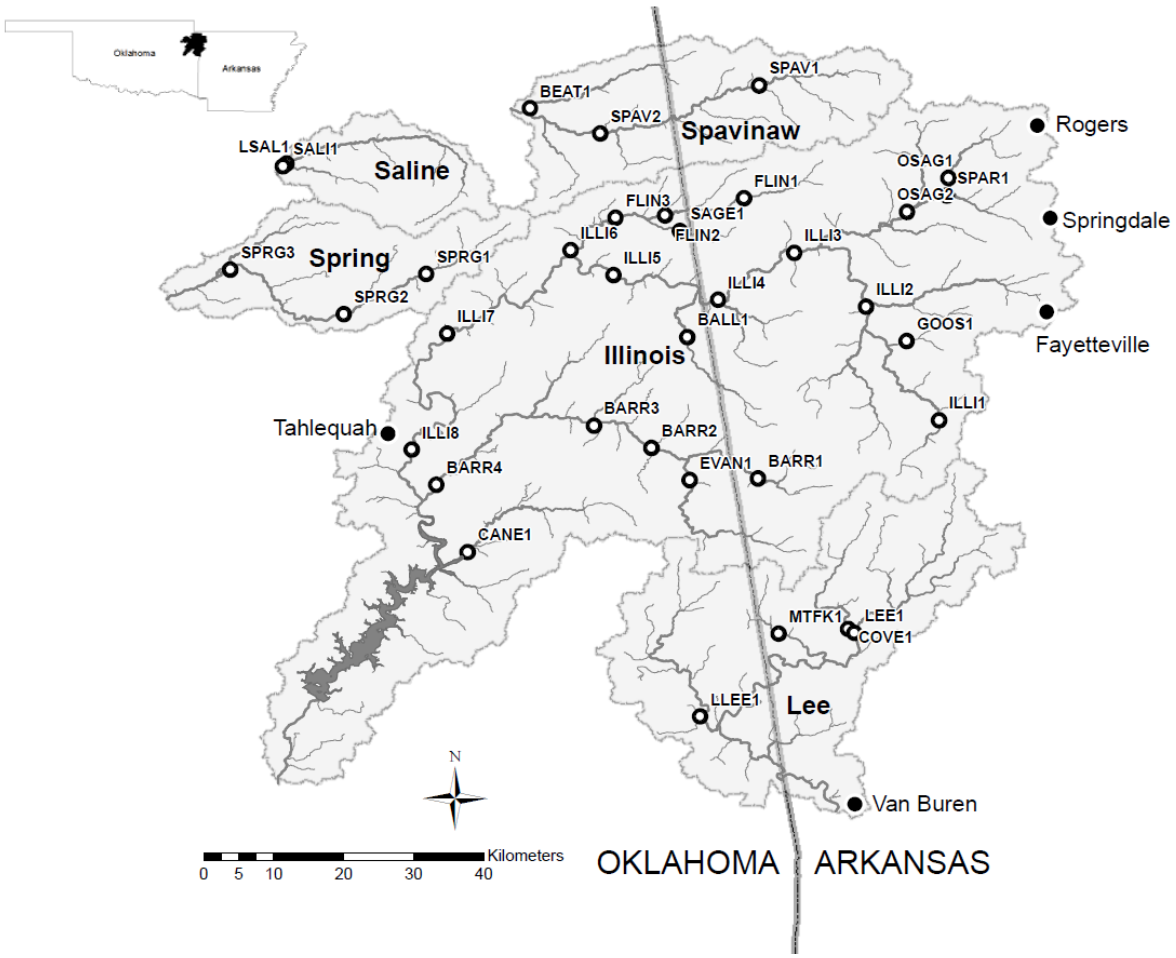

Figure 1. Locations and site codes of the 35 sampling reaches (see Table 1).

Catchment land cover/land use

Land cover and land use in the catchments of the 35 sites varied primarily in the percentage cover of forest, pasture, or developed land (Figure 2, Table 2). Most sites, even those with relatively low levels of total phophorus, had at least 30% cover of pasture land. The exceptions were COVE1, LEE1, LLEE1, and MTFK1, catchments that skirted the edge of the Ozark Highlands and were primarily located in the adjacent Boston Mountains. These sites had steeper uplands that limited extensive ranching and development. However, pasture land in the these catchments was typically located near the stream, where, if a source of phosphorus, may have a greater effect on nutrients than if located farther away (e.g., King et al. 2005). Moreover, these sites had similar levels of total phosphorus as sites with the lowest levels of pasture in the Ozark Highlands ecoregion (0.005-0.01 mg/L TP).

Sites that had relatively high levels of impervious cover associated with urban development were on the low end of urban intensity indices when compared to major metropolitan areas around the world (e.g., Walsh et al. 2005). Only 4 sites exceeded 10% impervious cover, and each of these were included because they had wastewater effluent discharges from sewage treatment plants upstream of our sampling reaches. Although levels of impervious cover exceeding 10% are known to have negative effects on benthic macroinvertebrate diversity (e.g., King et al. 2011), this may be less true in large streams and wadeable rivers such as those in our study, where the effects of imperviousness on storm runoff and peak flows is diminished.

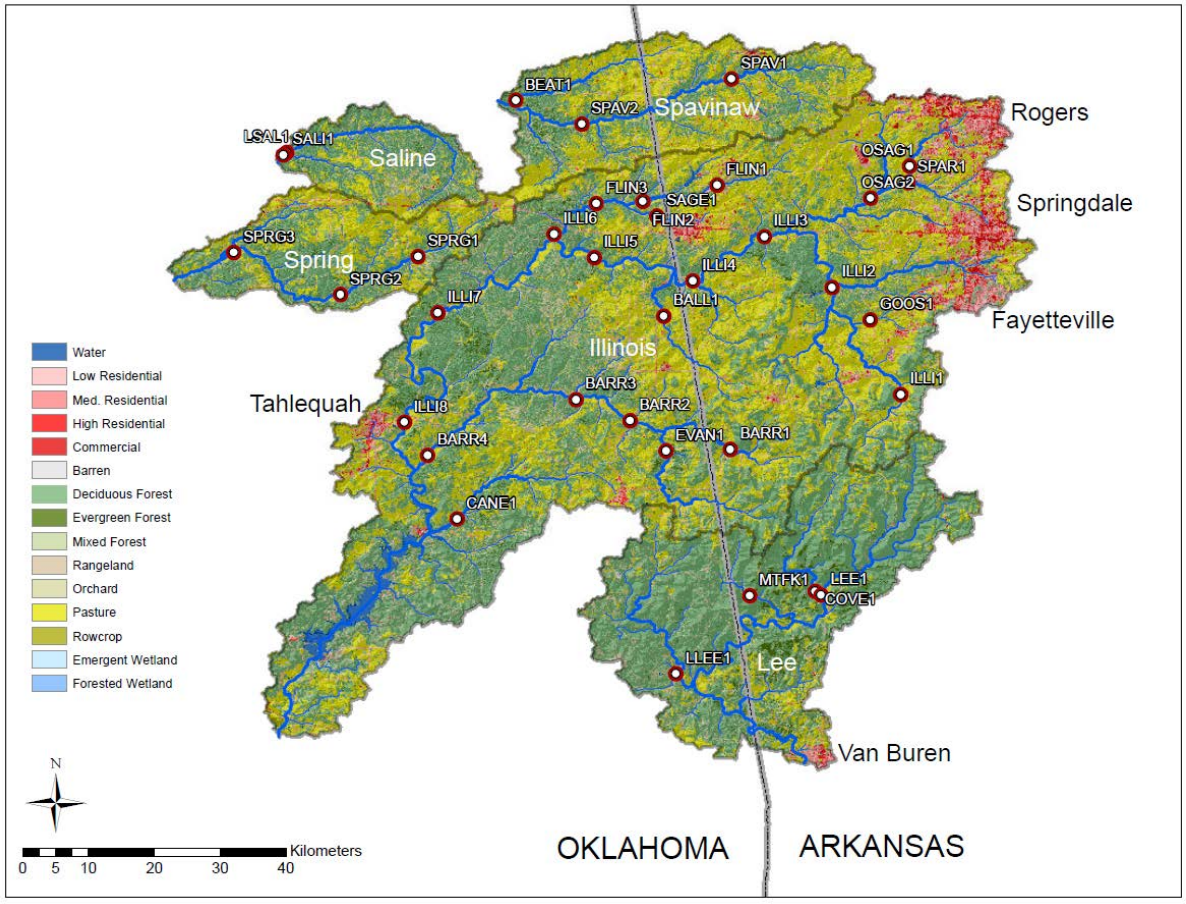

Figure 2. Land use and land cover patterns within the study area.

Table 2. Catchment area and percentages of dominant land cover classes associated with each sampling location. Land cover data was extracted from the most recent version of the National Land Cover Dataset (NLCD 2011).

Site ID	Catchment area (km^2)	% Developed	% Impervious cover	% Forest	% Grassland	% Pasture	% Row crop	% Wetland
BALL1	90.2	7.86	1.31	23.19	0.99	67.73	0.04	0.12
BARR1	105.6	4.22	0.57	45.43	1.82	48.28	0.00	0.14
BARR2	409.5	4.57	0.48	47.63	2.26	44.95	0.09	0.35
BARR3	542.9	4.97	0.56	46.06	2.90	45.37	0.08	0.34
BARR4	879.9	4.73	0.48	49.42	6.18	38.31	0.05	0.33
BEAT1	152.6	5.02	0.70	29.76	2.14	61.72	1.22	0.06
CANE1	232.9	5.93	0.98	43.54	3.37	46.71	0.10	0.09
COVE1	135.3	2.24	0.14	84.33	2.16	11.18	0.00	0.04
EVAN1	164.2	4.29	0.37	52.36	2.69	39.88	0.05	0.59
FLIN1	64.9	9.27	1.83	25.60	2.79	61.50	0.00	0.35
FLIN2	145.9	9.06	1.72	27.56	3.02	58.09	0.20	0.37
FLIN3	245.2	13.18	3.59	27.94	3.62	53.43	0.24	0.36
GOOS1	35.5	23.51	6.96	26.13	0.83	49.21	0.12	0.17
ILLI1	68.9	4.52	0.44	55.61	2.70	36.85	0.06	0.25
ILLI2	420.4	8.33	1.67	34.97	1.43	54.30	0.11	0.44
ILLI3	1239.8	20.84	6.53	27.11	1.16	49.75	0.12	0.42
ILLI4	1473.7	18.38	5.63	28.18	1.18	51.17	0.11	0.44
ILLI5	1716.9	16.85	5.00	29.09	1.25	51.70	0.12	0.48
ILLI6	2092.8	15.73	4.57	30.67	1.99	50.36	0.13	0.49
ILLI7	2294.6	14.64	4.18	34.05	2.84	46.99	0.12	0.55
ILLI8	2465.6	13.91	3.92	36.70	3.01	44.76	0.11	0.66
LEE1	252.2	2.73	0.24	84.62	2.17	9.93	0.01	0.27
LLEE1	264.1	2.79	0.16	77.98	8.53	9.22	0.00	0.19
LSAL1	61.7	3.31	0.33	50.93	8.32	34.89	0.43	0.00
MTFK1	67.1	2.45	0.10	84.70	4.94	7.02	0.00	0.03
OSAG1	100.8	56.47	21.50	7.27	0.37	34.57	0.20	0.20
OSAG2	337.4	36.94	13.02	11.29	0.36	50.38	0.16	0.15
SAGE1	45.9	35.50	12.99	8.99	1.18	53.63	0.03	0.23
SALI1	270.1	4.01	0.40	60.02	7.59	26.34	0.16	0.14
SPAR1	91.7	44.02	16.31	11.69	0.24	42.69	0.01	0.10
SPAV1	173.9	7.34	1.19	38.54	2.30	51.50	0.03	0.07
SPAV2	421.6	6.41	1.09	38.10	2.04	52.91	0.28	0.09
SPRG1	84.0	8.38	1.20	29.87	4.10	56.92	0.00	0.16
SPRG2	194.8	5.79	0.71	39.09	4.01	50.36	0.00	0.25
SPRG3	296.7	4.65	0.51	50.41	3.83	40.38	0.00	0.34

Sampling frequency

Sampling occurred on bimonthly schedule, subject to weather and stream flows. We chose to sample at this frequency for two years to increase the likelihood that we would detect nuisance algal blooms if they occurred (Biggs 2000). This sampling frequency resulted in 12 events (hereafter, Events 1-12), with 35 streams sampled per event, from June 2014 through April 2016 (Table 3), in addition to the total phosphorus (TP) data collected in April 2014 (hereafter, Event 0).

Table 3. Schedule of sampling events. Comprehensive sampling occurred bimonthly starting in June 2014 through April 2016, whereas total phosphorus sampling began in April 2014.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014				Site selection		X		X		X		X
2015		X		X		X		X		X		X
2016		X		X		Analyses, meetings, and final report completion						

Field Methods

Transect delineation

Field methods were patterned after Barbour et al. (1999) and Biggs and Kilroy (2000). Three transects were delineated to span a cross-section of each stream. Transects were delineated upon each site visit and did not necessarily correspond to previous transect locations because of different water levels or flood events that changed channel units between events.

For large streams/rivers (e.g., middle and lower Illinois River, lower Barren Fork Creek, Lee Creek, and several others), we typically identified a single riffle channel unit. The channel unit often was a large riffle that extended to deeper water, whereby three transects began at the wetted margin of the stream out to the point in the stream deemed representative of riffle-glide habitat or before it was too deep or fast to safely sample. The longitudinal distribution of these transects were roughly equidistant from the upper to lower boundaries of the riffle, but were always placed to target medium-large cobble (10-20 cm) habitat.

For streams with riffles that were wadeable from bank to bank and had a series of riffle-pool channel units within a relatively short length of longitudinal reach (<100 m), we selected 3 riffle channel units and placed one transect in each unit. Transects spanned the width of the optimal habitat, which typically was equal to the wetted width of the stream but occasionally was truncated by a pool, a change in substrate, heavy shade, etc, along one margin of the stream. Here, transects extended from one bank out to the margin of the cross-section that had the appropriate depth, velocity, light, and substrate.

Five sampling points were marked along each transect, roughly equidistant but allowing for some variability in location to ensure appropriate depth, velocity, light, and substrate. The first and last points were within 1-2 m from each end transects. Points 2, 3, and 4 were marked at 0.25, 0.5 and 0.75 distances of transects. Points were marked on the stream bottom using flagging tape secured to a large, galvanized metal washer.

Surface water chemistry and phytoplankton collection

Water chemistry and seston samples were collected above the upstream boundary of the reach after the upstream transect was marked. Triplicate TP samples were collected in new 50 mL centrifuge tubes and immediately preserved with sufficient volume of H_2SO_4 to achieve pH < 2. A single grab sample per site was collected for each of the following: TN (unfiltered, preserved with H_2SO_4) and NH_4-N, NO_2+NO_3-N, and PO_4-P (field filtered, 0.45 µm, iced immediately, held at <4 C until frozen that evening.). Separate 1-L sestonic chlorophyll-a and total suspended solid samples were collected in dark bottles and placed on ice immediately. Sample collection followed the Baylor University Center for Reservoir and Aquatic Systems Research (CRASR) approved quality assurance/quality control protocols.

Site characterization

We measured the following physical and chemical variables to characterize the reach on every visit: wetted width (estimated when wading the full width of the stream was not possible), mean depth (m) and velocity (m/s) of riffle channel unit (corresponding to benthic algal sampling transects), canopy cover (0-100%), discharge (ft^3/s, and several conventional water quality variables.

Discharge was estimated using a Marsh-McBirney flowmeter following standard USGS protocols. Discharge generally was not measured at sites that were (a) gaged and had moderate to high flow at the time of sampling, and (b) too large or unsafe wade (mainstem Illinois River). Discharge at gaged sites was estimated during summer low-flow conditions if it can be accomplished safely.

Temperature, specific conductivity, pH and optical dissolved oxygen were measured using YSI EXO1 multiprobes deployed for a minimum of 15 minutes during the site visit. Multiprobes were placed in flowing water above the reach. Readings were recorded manually after sensor readings stabilized. Multiprobes were calibrated prior to each event and post-calibration checked following each event.

Periphyton collection

Cobbles were collected at each of 15 points starting with the most downstream transect. The cobble nearest the transect marker that was 10-20 cm wide was selected regardless of the amount of algae on the top of the substrate, although oil shale fragments were excluded from sampling because they were rare. Rather, calcite or dolomite, the two dominant rock types in these streams, were selected.

Cobbles were removed from the stream by carefully lifting the substrate slowly to the surface. Each substrate was carefully placed in a white sampling basin designated for that transect. This process was repeated until cobbles from each of the 5 points were collected, and repeated again for each of the 2 remaining transects.

Each white basin was partially filled with stream water to keep the periphyton from desiccating and for enhancing the quality of photographs. Each white basin was photographed separately prior to removal of attached periphyton. A small white board with the date, site and transect ID, and event number marked using a dry erase marker was included in each photo to assist with cataloging of photos.

Periphyton was removed from the 15 cobbles before leaving the site. Cobbles were scraped over a clean, deep-sided white pan using a stainless steel wire brush. All attached algae was removed from the upper surface of the cobble. Stream water was used to rinse residue from the cobble into the white pan. After all cobbles were scraped and rinsed, the contents were consolidated into one corner of the pan and poured into a 1 L dark bottle, which was immediately placed on ice to achieve a sample temperature of < 4 degrees C until processing later that day.

Following the removal of periphyton from cobbles, the upper surface of each cobble was wrapped with aluminum foil for estimating the area (cm^2) from which the periphyton was removed. Foil was carefully cut along the margins of the cobble corresponding to the perimeter of the area sampled, removed, and placed in a labeled bag. This process was repeated for all 15 cobbles prior to leaving the site. Foil was cleaned, dried and weighed using analytical balance. Total mass of foil per site was used to estimate area using a simple weight-to-area conversion factor.

Hess (macroinvertebrate) sampling and transect marker characterization

Macroinvertebrate sampling was done primarily to estimate the density and biomass of periphyton grazing taxa, particularly snails in the family Pleuroceridae. Grazing taxa can achieve high densities and exert strong top-down control on algal biomass, hence quantifying their abundance was considered an important ancillary measurement to help explain patterns of benthic algal biomass over time.

Quantitative macroinvertebrate samples were collected using a Hess sampler approximately 0.5 m upstream of each of the 15 transect markers. The Hess sampler was placed upstream to avoid where the periphyton cobble was collected or where anyone had walked or otherwise disrupted the substrate.

Once the Hess sampler was embedded into the substrate, water depth, dominant substrate (gravel or cobble), sedimentation index (qualitative, 1-20, similar to EPA RBP; Barbour et al. 1999), embeddedness of cobbles (0-100%), and stoneroller grazing scars (qualitative, 0-10) within the Hess sampler was recorded prior to disruption of the substrate in the sampler. Next, all gravel and cobble were thoroughly brushed to remove attached periphyton, organic matter, and aquatic macroinvertebrates. Brushing was done inside the sampler where material and organisms were flushed back into the trailing net. Once all surface rocks had been brushed and removed, the remaining substrate was vigorously agitated to a depth of 5 cm for at least 30 seconds to dislodge remaining organisms. Following this step, the Hess sampler was carefully but quickly lifted off of the bottom to help rinse material attached to the net into the dolphin bucket attached to the cod end of the net. Additional rinsing of material from the net into the dolphin bucket was done as necessary. Contents of the dolphin bucket were emptied into a heavy-duty plastic 4-L storage, which was eventually used to composite all 15 Hess samples from one site. Additional storage bags were used if necessary. Before leaving the site, the sample bag(s) was placed on ice for preservation using buffered formal at the temporary field lab later that same day. The final volume-to-volume concentration of formalin after being mixed with the sample material in the bag met or exceeded 5%.

Diel dissolved oxygen and pH

We deployed YSI EXO1 data sondes to measure optical dissolved oxygen (DO) and pH at 15-minute intervals for approximately 48 h at a minimum of 25 sites in summer 2014 and 2015.

The purpose of measuring diel variability in these water quality variables was to determine whether TP was correlated with minimum dissolved oxygen and maximum pH. Both variables are mechanistically related to primary production in streams, but also are strongly influenced by differences in water turbulence (reaeration) among sites, groundwater discharge in the reach, and light conditions during deployment, all of which are very difficult to account for in the large streams and rivers sampled in this study.

Sondes were deployed at a depth of approximately 0.5 m. Sondes were located in shallow glide-pool habitats above riffles in order to reduce the effect of reaeration on DO and pH. Sondes were calibrated immediately prior to deployment, and post-calibration checks were performed following deployment. Sondes that failed post-calibration were excluded from analysis, as were sondes that were affected by factors that biased the results, such as accumulation of drifting debris (which was noted upon retrieval) or an obvious groundwater input immediately adjacent to the deployment site (which was discovered upon reviewing the data).

Frequency and Duration of Stressor and Response Variables

Two critical elements of developing a numerical criterion for total phosphorus for the Designated Scenic Rivers are sampling frequency (how often a TP sample is collected) and duration (over what period of time is the numerical criterion assessed, averaged, and evaluated for exceedance). A third element is frequency of excursion during a defined assessment period to meet the criterion, but this beyond the scope of this report.

Sampling frequency in our study was established during the study design phase prior to collection of any samples. Samples were collected bimonthly during base flow conditions only (or "critical flow" as defined by the Joint Study Committee, which were any flow conditions that were not dominated by surface-water runoff). The decision to sample during base flow conditions was based on several key factors: (1) it was impractical if not impossible under this budget to collect nearly continuous (daily to multiple times per day) samples to estimate phosphorus concentrations representative of all flow conditions from 35 locations over a 2 year period, (2) base flow conditions provide a more representative estimate of phosphorus availability to benthic algae because storm flows usually result in scouring of algae from rocks and very high turbidity which is not conducive for algal growth due to attenuation of light, (3) base flows occur the vast majority of the time, thus they are the typical condition in streams, (4) US EPA recommends and many other states use base flow conditions to establish numerical criteria for streams and rivers, thus there is a precedent for using data collected only during base flow for estimating violations of a numerical criterion, and (5) base flow TP is typically strongly correlated to TP calculated across all flow conditions where such data are available (e.g., Figure 3)

Duration was constrained by the length of the study (2 y) and was assessed by the comparing the strength of the relationships between mean TP calculated across different time intervals to biological response variables, particularly algal biomass. Mean TP (mg/L) was calculated at 2, 4, 6, 8, 10, and 12 month intervals. A 2-month interval included TP samples from 2 events; for example, our first algal sampling event was June 2014, whereas our first phosphorus sampling

event was April 2014 (Event 0). The mean of April and June 2014 TP was the value used when relating 2 month TP to benthic chlorophyll-a collected in June 2014 (see Data Analysis). Similarly, 4 month TP was calculated as the mean of the 2 previous events and the current event (e.g., Events 0, 1, and 2), and so forth. We used arithmetic mean because it was almost perfectly correlated to geometric mean (Figure 4) and is likely a better estimator of cumulative exposure.

Response variables were analyzed as instantaneous measurements (e.g., 4 month mean TP vs. the observed level of benthic chlorophyll-a on a particular event that matched the 4 month TP window) and as mean responses that matched the TP duration (e.g., 4 month mean TP vs. the mean of benthic chlorophyll-a matching the same events used to calculate the 4 month TP; Figure 5). US EPA (2010) recommended calculating mean nutrient and response data if multiple collections were available from the same locations over time because it reduces variability, improves statistical models, and is consistent with the way numerical criteria are assessed (typically over a series of months or a year or more).

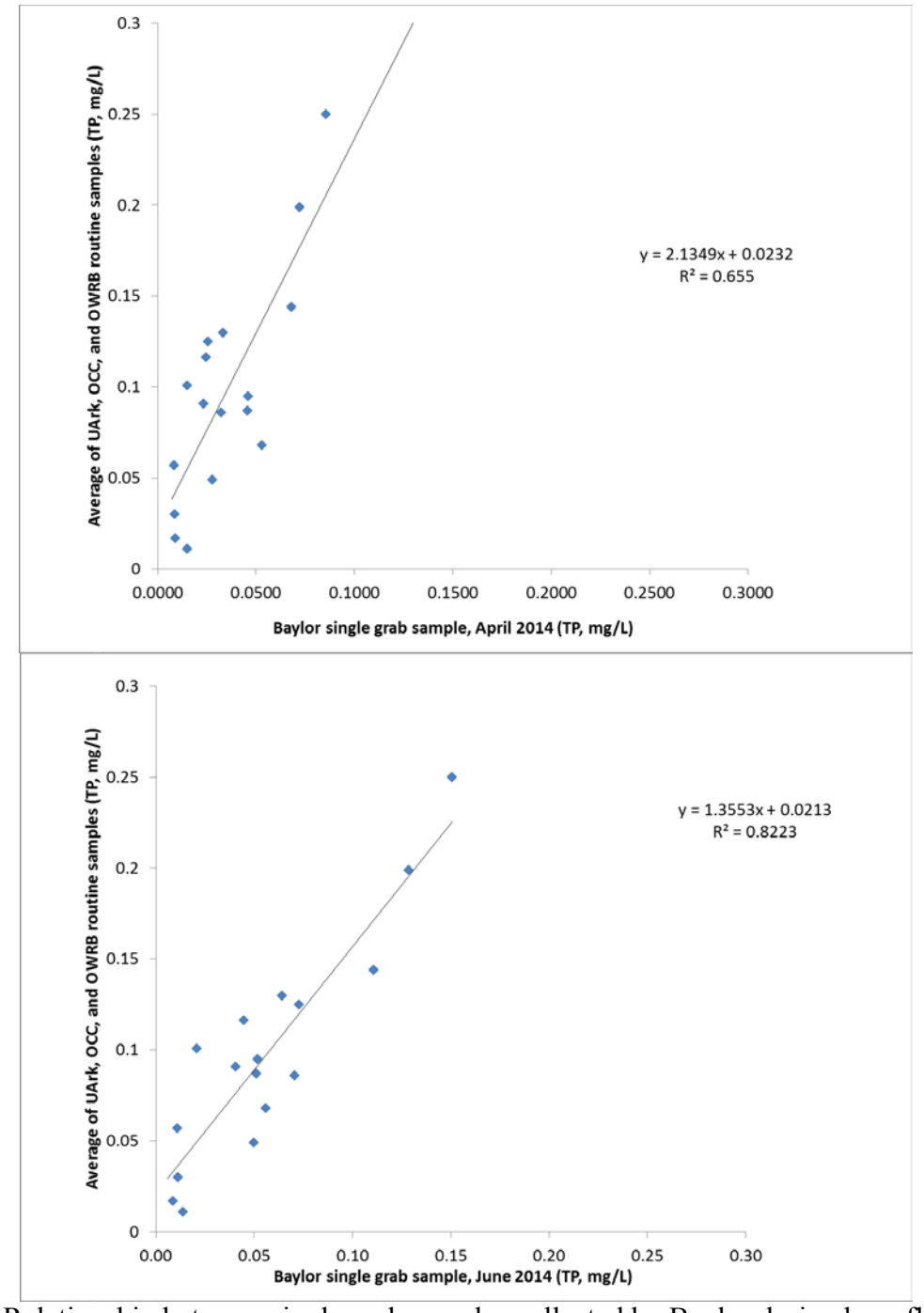

Figure 3. Relationship between single grab samples collected by Baylor during base flow in April (upper panel) and June (lower panel) 2014 to mean TP over 1-2 years prior to the Baylor samples from intensive sites (i.e., samples collected at any flow, including storm flows) monitored by the Oklahoma Conservation Commission (OCC), Oklahoma Water Resources Board (OWRB), and the University of Arkansas (UA).

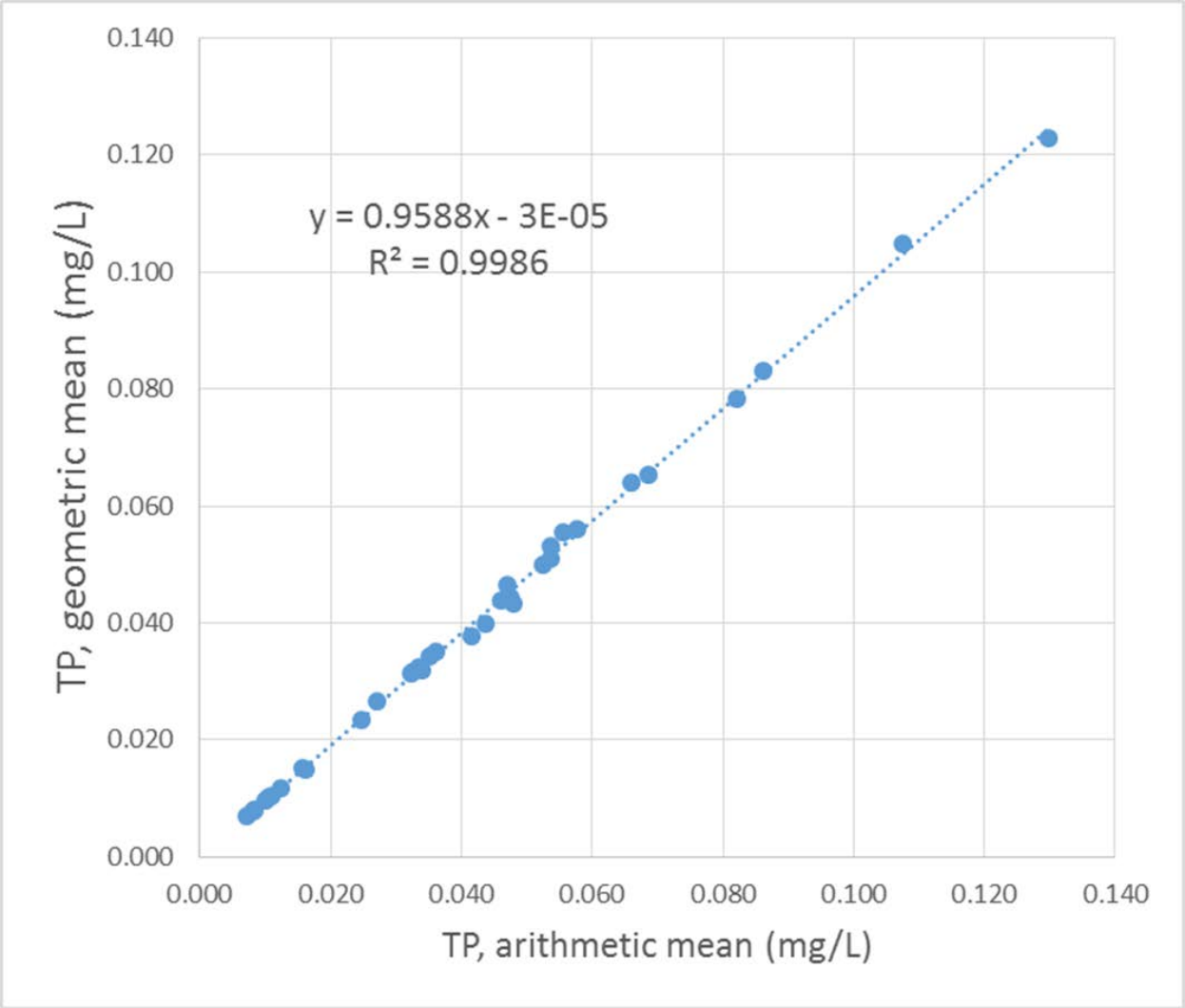

Figure 4. The relationship between geometric and arithmetic mean total phosphorus concentrations from the 35 study sites from April 2014 through April 2016 (n=13).

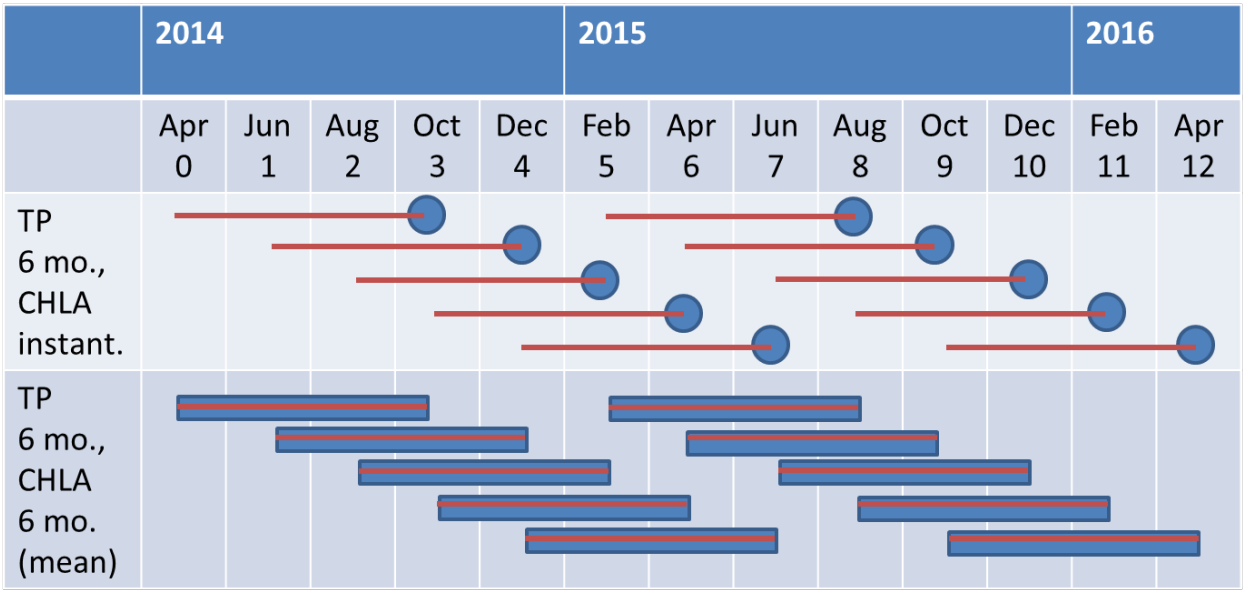

Figure 5. Examples of the two different ways total phosphorus was related to biological response variables. This example is based on a 6 month mean TP. In the top row, each dot represents the "instantaneous" set of values of benthic chlorophyll-a measured on each of the events, and the red line represents the time interval (duration) over which TP was averaged prior to relating to these instantaneous measures of chlorophyll-a. In the bottom row, the blue bars represent the time interval used to calculate the "mean" set of response values of benthic chlorophyll-a, which matches the same set of data used to calculate the mean TP.

Data Analysis

The primary purpose of the Scenic Rivers Joint Phosphorus Study, as stated by the Second Statement of Joint Principals and Actions, page 2, was to identify "*the total phosphorus threshold response level....at which any statistically significant shift occurs in algal species composition or algal biomass production...resulting in undesirable aesthetic or water quality...conditions in the Designated Scenic Rivers."*

A threshold level of TP, defined ecologically, is a where there is a disproportionately large change in an ecological response, such as algal biomass or species composition, with a relatively small incremental increase in concentration of TP (Groffman et al. 2003, Baker and King 2010).

Statistically, a stressor-response threshold can be categorized into two broad, but complementary classes of methods. The first, a change point threshold approach, relates to finding value along a stressor gradient where the response variable, such as algal biomass, changes the most. Here, the goal is to estimate the level of the stressor (the x axis, or predictor variable) where the mean of a response variable increases or decreases disproportionately, such that by splitting the data into two groups defined as above and below that point, the means of those two groups would differ the most when compared to all other possible values of TP in the data set (Figure 6).

The second approach involves identifying the value of the predictor where the mean (or median or other quantile) of the response (the fitted line of a regression, for example) intersects a critical reference value of the response, such as a minimum dissolved oxygen or nuisance levels of benthic chlorophyll-a (Figure 7). This reference value approach is ideal for a policy-based study where an *a priori* management target or standard has been previously established. The first approach is very useful when a management target is not defined or there is an additional goal of identifying where there is the largest change, regardless of a management target (e.g., *"any statistically significant shift occurs…."*, p2, Second Statement of Joint Principles and Actions). However, it should be made clear that the first approach is based on splitting the data at the point of greatest change, but "greatest change" may not correspond to a reference value threshold for a particular endpoint. Here, we describe both approaches and how they were used to satisfy the primary purpose of the Second Statement of Joint Principles and Actions.

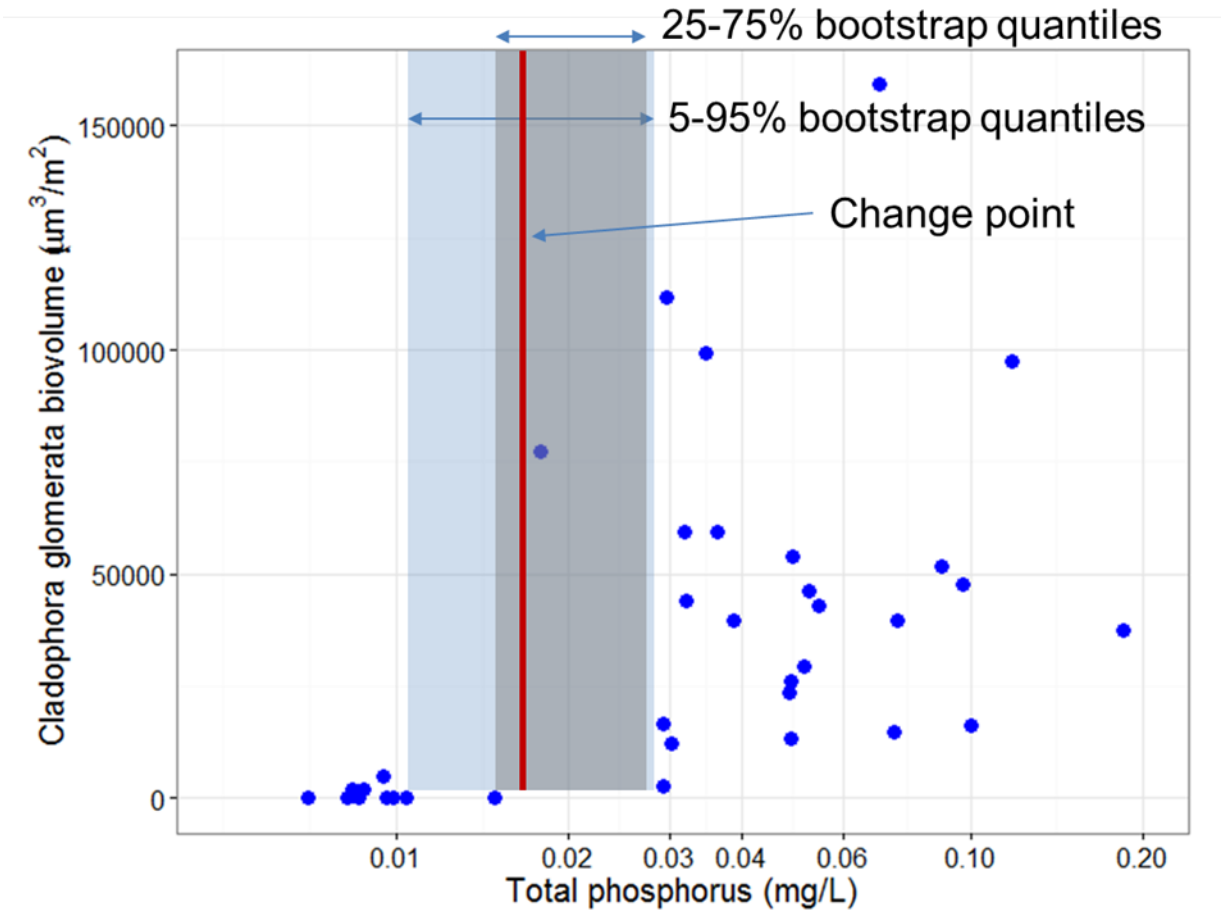

Figure 6. Change point threshold approach based on splitting the data at a TP value that corresponds to the largest change in the response (in this case, biovolume of *Cladophora,* the primary nuisance species in the Designated Scenic Rivers). Here, the data are 2 month TP versus instantaneous *Cladophora* biovolume from June 2014.

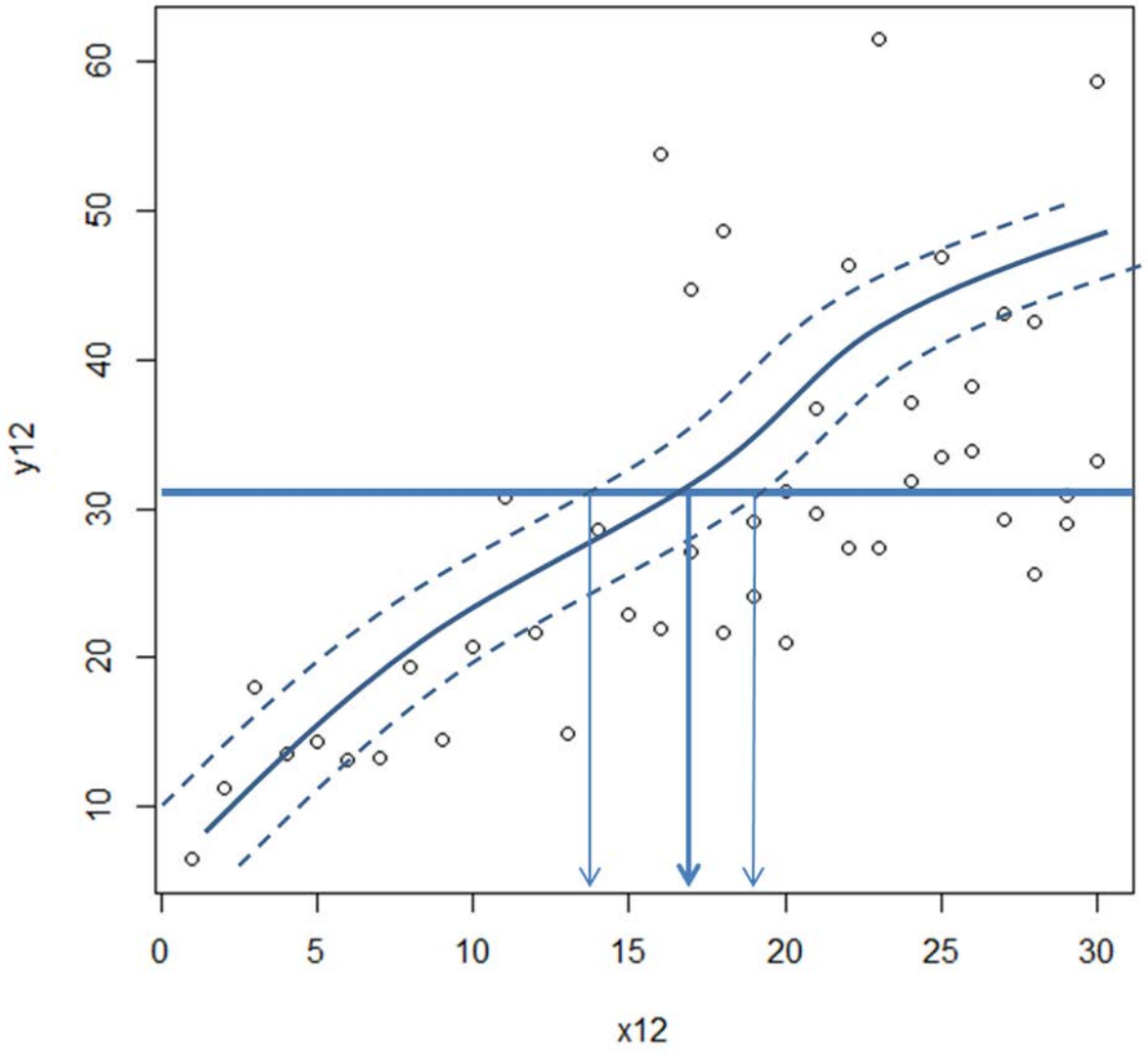

Figure 7. An illustration of the reference value approach. Here, the theoretical reference value is 30, which presumably represents a biological criterion beyond which conditions are considered unacceptable. The fitted line and confidence limits (dotted lines) are used to statistically estimate the level of the predictor (labeled "x12" in this example) that results in an intersection with the y-axis reference value. Here, the mean fitted response intersects the reference value at an x-axis value of approximately 17, whereas the lower and upper confidence limits intersect the reference value at 14 and 19, respectively. Thus, levels of the stressor (x12) that exceed 17, with uncertainty of 14-19, are likely to violate the biological response reference value (y12=30).

Change-point threshold approaches

Nonparametric change-point analysis

There are several methods for estimating statistical change points, but many are not well suited for ecological data (see list of methods in Dodds et al. 2010). A nonparametric form of change point analysis that was employed by King and Richardson (2003) and is included as a recommended technique for deriving numeric nutrient criteria by US EPA (2010) is one of the few techniques that makes few implicit assumptions about the data, particularly ones almost always violated by comparable methods (e.g., piecewise linear regression), despite their widespread use (e.g., Toms and Lesperance 2003). Nonparametric change point analysis, or nCPA as implemented in King and Richardson (2003), is simply a restricted form of regression tree analysis (De'ath et al. 2002) that involves only one predictor and one "branch" in the tree. The branches are defined by the change point. However, there are a few important limitations of using a simple regression tree to identify change points.

First, regression tree analysis identifies one value of the predictor (in this case, TP) that results in the greatest amount of variance explained (more technically, deviance), yet many other values of the predictor may explain very similar amounts of variance. In many stressor-response relationships, there is a zone of disproportionate change (see the gray area in Figure 3) where any one of several values in a relatively narrow range are nearly interchangeable in their ability to explain the variance in the response. To deal with this limitation, the change point approaches employed in this report use a bootstrapping algorithm to estimate quantile intervals (similar to confidence intervals) that provide estimates of uncertainty about where the true change point might be located, if there is one. This is very similar to the use of bootstrapping in Random Forest analysis, a related technique (Breiman 2001).

Second, most simple regression tree analyses do not include an estimate of statistical significance, and those that do often assume a normal distribution, which is inappropriate. The nCPA method employed in this report uses a randomization test to estimate the probability that the variance explained by the model is not better than expected by chance, with a minimum of 1000 randomizations.

Third, the version of nCPA used in this study employs several different probability distributions for calculating deviance reduction (Gaussian, binomial, Poisson) depending upon the type of response data. For example, the proportion of biovolume as nuisance algae species is a binomial response variable and thus a binomial form of nCPA was employed for that analysis.

Change point analysis has its own share of limitations, however. First, the analysis can yield biased change point estimates if the predictor data is strongly skewed (i.e., many high values and very few low, or vice-versa). However, this is a problem for all statistical methods and is a particular problem in observational stressor-response studies that are not carefully designed to sample a stressor gradient in a relatively uniform manner (King and Baker 2014). Second, the method will find a change point even if the response to the predictor is a linear relationship because there is significant change associated with a linear relationship. However, the bootstrapping method largely alleviates this concern because the quantile intervals will span most of the range of x, indicating that the point of greatest change is highly uncertain and could

be almost anywhere along the gradient. Thus, using the bootstrap results in conjunction with common sense (i.e., visualizing the data using scatterplots prior to conducting the analysis, e.g. Zuur et al. 2010) allows for strong inferences to be made.

In accordance with recommendations by the SRJSC, change-point analysis was used to estimate TP change-points for the following variables: algal biomass (benthic chlorophyll-a), *Cladophora* biovolume, and the proportion of nuisance algal taxa, the three primary variables of interest for assessing the relationship between TP and nuisance levels of algal biomass.

Threshold Indicator Taxa Analysis (TITAN)

TITAN (Baker and King 2010) is an analytical approach for identifying and distinguishing threshold-type responses among many species simultaneously in response to a stressor gradient (e.g., algal species composition). King and Baker (2014) provide explicit detail on its use, misuse, and limitations for natural resource management. Briefly TITAN works by integrating a relatively simple and elegant measure of association in taxon abundance with a nonparametric technique for detecting change. Indicator species analysis (Dufrene and Legendre 1997) uses abundance-weighted occurrence frequency to describe association between a particular taxon and groups of samples defined by their order along an environmental gradient. To facilitate comparison across taxa, TITAN compares each taxon's maximum IndVal score to those expected if the same sampled abundances were randomly distributed across the environmental gradient. A good indicator species is one that occurs frequently at one end of a gradient, so that changes in its abundance are easy to detect, but that is not the only kind of response worth noting. IndVal scores will always be small for rare, variable, or sensitive taxa, even though they can nonetheless represent important changes within a community. By comparison to the average IndVal scores derived by random permutation, TITAN standardizes measures of change for any given taxon to units of standard deviation (z scores; Baker and King 2010). Standardization emphasizes observed changes for each taxon relative to their own patterns of variability in abundance and occurrence.

To better understand uncertainty surrounding the observed change points, TITAN employs a bootstrap resampling technique in the same way the previously described nCPA method does. Information provided by the bootstrap is critical for interpreting results in TITAN. In addition to estimation of change-point quantiles, TITAN evaluates consistency in the response direction as purity, and the frequency of a strong response magnitude as reliability (Baker and King 2010). Combined with a minimum occurrence frequency, these diagnostic indices are used as filters to help distinguish the signal produced by indicator taxa responses from stochastic noise along the gradient. This filtering is part of what distinguishes TITAN from many other multivariate techniques based on weighted averaging or dissimilarity.

Once indicator taxa have been identified, TITAN provides information that can be used to identify a potential community-level threshold. A plot of filtered indicator taxa showing change-point quantiles from bootstrap replicates provides evidence regarding the existence of synchronous changes in the community structure (Figure 8, Texas stream example). Because the magnitude of all responses is standardized across taxa as z scores, their sum reflects the magnitude of community change at any point along the gradient. Distinct peaks in the sum(z)

curve (maxima) plotted across the environmental gradient are another indication of coincident change in community structure. When bootstrap replicates used to compare the location of the sum(z) maxima across many sample replicates show a narrow band, this constitutes evidence for a threshold response (Baker and King 2010; King et al. 2011).

TITAN was used to estimate taxa-specific change points and community-level thresholds in algal species abundance (biovolume/cm^2) in response to TP.

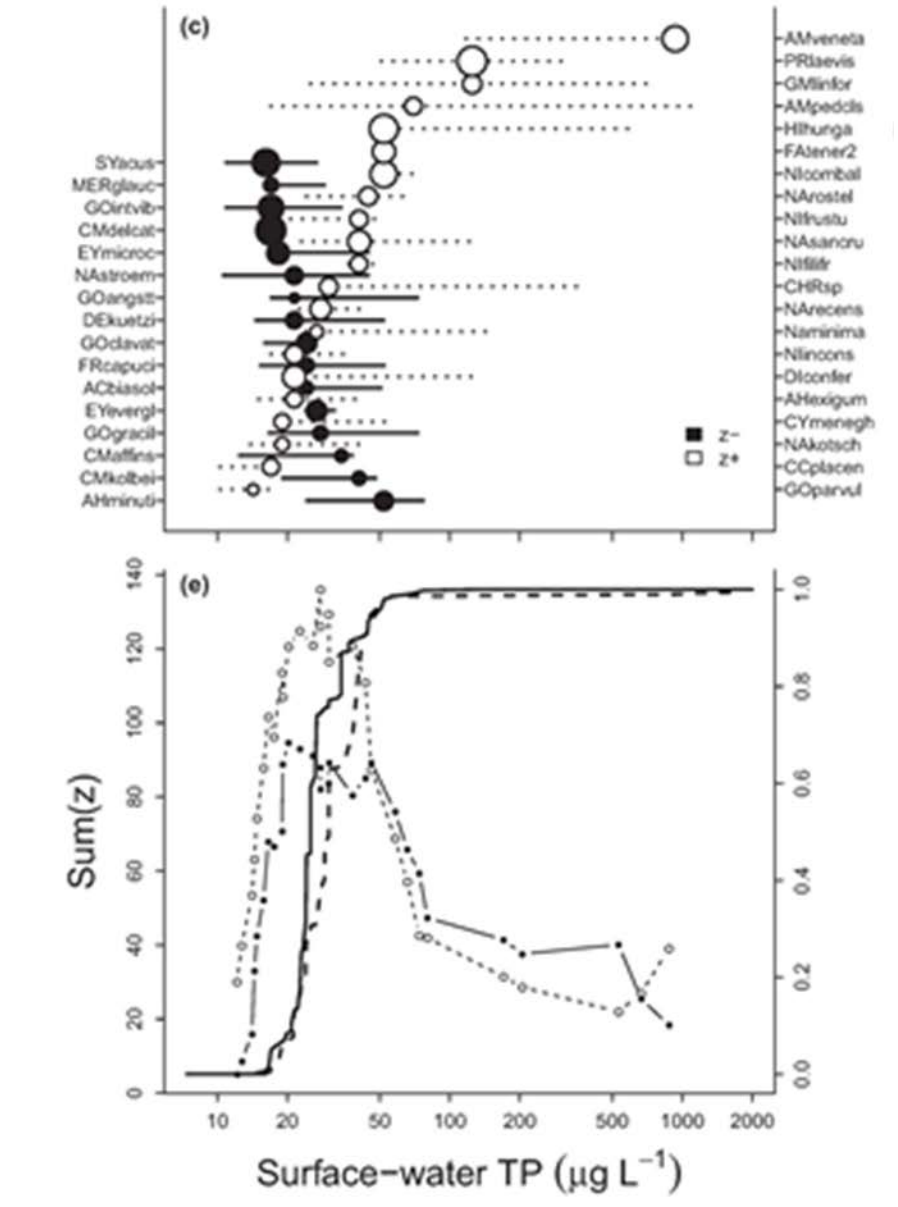

Figure 8. Example of output from Threshold Indicator Taxa Analysis (TITAN). In this example from a study conducted in wadeable streams in central Texas (Taylor et al. 2014), species with negative responses to total phosphorus are shown as filled symbols, whereas species that increased in response to TP are shown as open circles (upper panel). The location of the symbols corresponds to the level of TP resulting the greatest change in the frequency and abundance of each taxon (the change point) and the horizontal lines span the lower to upper quantile intervals (uncertainty). The lower panel illustrates the sum of the responses of the pure and reliable threshold indicator taxa. Sum(z-) (negative responding taxa) sharply peaks at 0.021 mg/L TP with lower and upper quantile limits of 0.016-0.052 mg/L. Sum(z+) (positive responding taxa) sharply peaked at 0.028 (0.018-0.048) mg/L TP. Both results are indicative of a significant shift in species composition between ~0.02-0.05 mg/L TP.

Reference value threshold approach

Neither Oklahoma nor Arkansas has numerical standards for benthic algal biomass or species composition. Scientific literature and a few states (e.g., Montana, Suplee et al. 2009) have either recommended or adopted ~150-200 mg/m^2 benthic chlorophyll-a as a management threshold, such that levels above this value represent nuisance levels of algal biomass. Thus, values of benthic chlorophyll-a at or above 150-200 mg/m^2 could be used as a reference values in this study for use in analyses that are set up to ask "at what level of TP does benthic chlorophyll-a exceed x mg/m^2?". However, differences between large streams and rivers in this study and those from typically much smaller streams in other regions of the world where these numbers have been adopted must be considered prior to using these reference values. Further, differences in taxonomic structure of periphyton in pristine streams of this region relative to other regions where those numbers have been adopted could result in lower or higher natural levels of benthic chlorophyll-a.

For these reasons, we examined values of benthic chlorophyll-a at sites at the low end of the TP gradient to assess the natural range of conditions that might be expected at reference sites in the Ozark Highlands and Boston Mountains ecoregions. Second, we fit an empirical relationships between benthic chlorophyll-a and biovolume of the dominant nuisance algal species in these streams, *Cladophora glomerata*, to refine estimates of nuisance levels of benthic algal biomass that were calibrated to these waterbodies (see Results for greater details).

Based on these assessments, we identified 150, 200, 250, and 300 mg/m^2 benthic chlorophyll-a as reference values representing potential nuisance levels of algae for the Designated Scenic Rivers. We assessed these reference levels using two methods.

First, we related mean benthic chlorophyll-a to year 1, year 2, and years 1 and 2 combined mean TP using a generalized additive modeling approach (GAM; Zuur 2009). A GAM model was the most appropriate for these response data because of nonlinearity that did not match a functional relationship (e.g., power, log, exponential). We used a Gamma probability distribution with an identity link function because the variance in the response was highly correlated to the predictor. Further, we weighted each mean by the inverse of its standard deviation (1/sd) so that points with higher variance associated with their means (more uncertainty) received less weight in the model.

Second, we analyzed the frequency of exceedance of each of those values as response variables to year 1, year 2, and years 1 and 2 combined mean TP using generalized linear models (GLM; Zuur 2009). We calculated the number of times each site exceeded 150, 200, 250, and 300 mg/m^2 benthic chlorophyll-a and fit a model based on a binomial (logistic) probability distribution to the data. The proportion of the total number of events per site in which benthic chlorophyll-a exceeded each of these values (4 separate response variables) was used as a response to mean TP. The total number of events, which was 12 for all but 3 sites that were either not flowing (ILLI1 and EVAN1, October 2014) or flooded (CANE1, June and December 2015) during our sampling event, was used as the weight for the binomial model (Zuur et al. 2009). The resulting models generated fitted responses of the proportion of times in which

benthic chlorophyll-a exceeded each of those 4 critical values for all levels of mean TP in the study.

Results

Temporal patterns in stream discharge, nutrients, and algal biomass

Sampling was successfully completed every two months during critical flow conditions at all of the 35 sites over the 2 year study, with the exception of two sites in October 2014 (ILLI1, EVAN1; streams were not flowing) and another site during June 2015 and December 2015 (CANE1; site was flooded by backwater from Lake Tenkiller).

Hydrographs (Figure 9) illustrate that 2014 through early 2015 was largely devoid of major storm flows associated with large precipitation events. This was not a particularly dry period, either, as precipitation was normal and base flows remained near the historical median for gaged sites. By April 2015, a much wetter weather pattern associated with El Niño conditions developed for the rest of the year, resulting in frequent storm flow conditions and culminating in an historic flood in late December 2015. The period following the historic flood was relatively dry and allowed the streams to return to high critical flow conditions by early February and relatively normal stream levels through March and April 2016.

Total phosphorus concentrations were relatively consistent within each stream over time with the exception of SAGE1, which was wastewater effluent dominated, and several other sites during periods of high primary production associated with blooms of *Cladophora glomerata*. In the latter instances, uptake by benthic algae reduced TP to levels 0.01-0.04 mg/L below the median TP value at these sites over the 2-y study (Figure 10). The patterns of benthic chlorophyll-a in this figure (symbols sized in proportion to chlorophyll-a values) also corroborate a very consistent pattern of sharp declines in TP with high levels of benthic chlorophyll-a.

Although not necessarily a focus of this study, it is important to acknowledge that nitrogen is also critical to primary production in streams, and has been suggested as possibly a stronger correlate of benthic chlorophyll-a in Ozark Highland streams in Arkansas. Because sources of phosphorus are almost always sources of nitrogen, too (e.g., wastewater discharges), it is logical that nitrogen should correlate well with benthic chlorophyll if phosphorus is also a good correlate. The problem with using simple correlations to ascribe causation is demonstrated, in part, in Figure 10 because it shows that during periods of high primary production, phosphorus is rapidly removed from the water column such that the relationship between TP and benthic chlorophyll-a at the particular point in time was weak, and probably weaker than the relationship to total nitrogen if nitrogen is not removed at the same rate as phosphorus, and particularly if it does not change relative to typical concentrations at that site.

To illustrate this point further, we plotted TP as the difference (deviation) from the median value measured at each site during the 2 year study (Figure 11). Large, negative deviations were almost always associated with disproportionately high levels of benthic chlorophyll and increasingly high N:P ratios, typically > 100 (Figure 11). Thus, it was the antecedent TP conditions that led to blooms, and when blooms were present, TP was being taken up more rapidly than it was desorbing from sediment or being supplied by wastewater (Figure 12). Conversely, TN showed no temporal pattern that related to benthic chlorophyll-a (Figure 12). Thus, this study's focus on P as the primary driver of potential nuisance conditions of algal biomass is well supported.

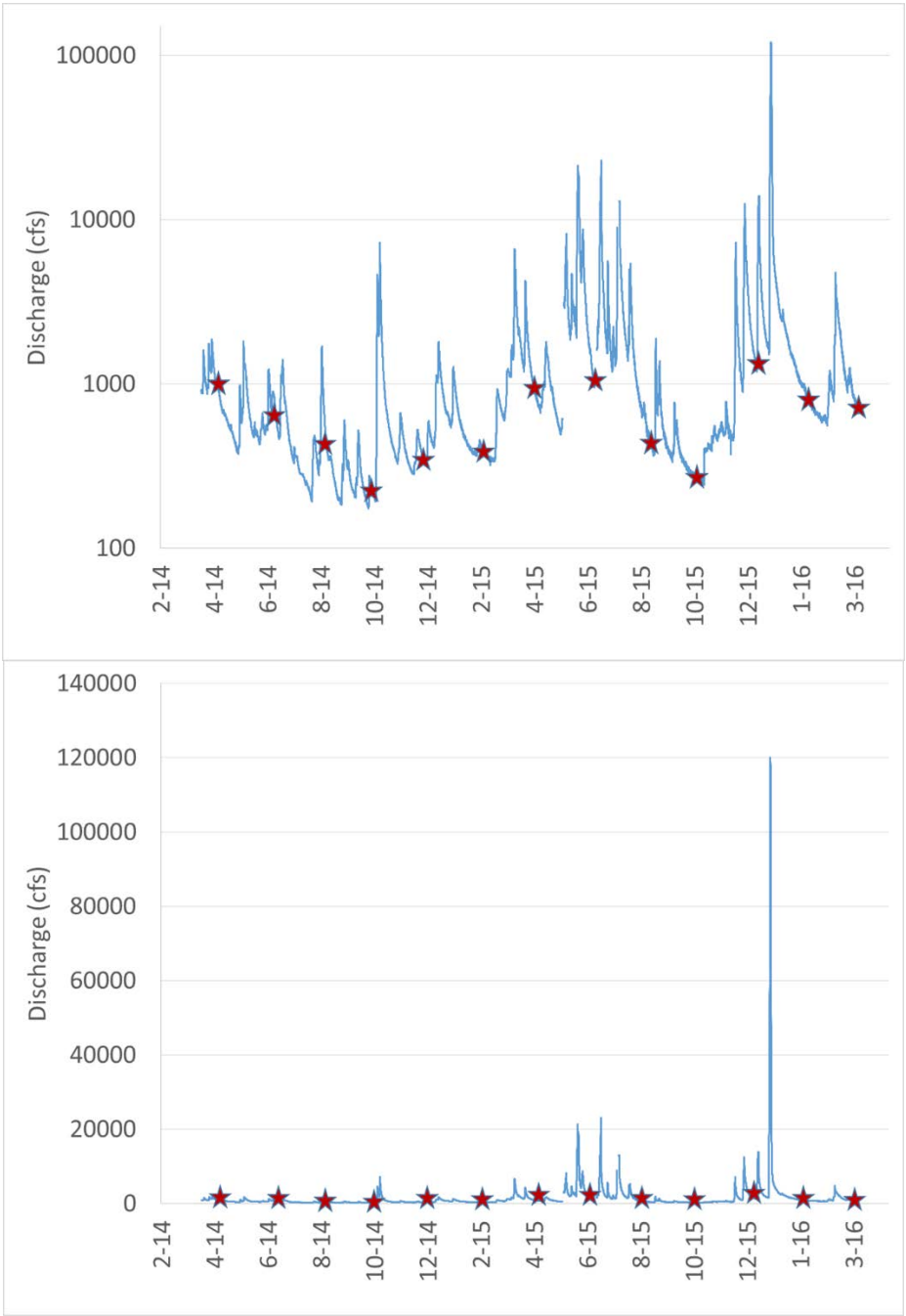

Figure 9. Daily mean discharge at USGS gage 07196500, Illinois River at Tahlequah, from April 2014-2016. Location of the stars indicates the approximate timing of sampling. Discharge is log-scaled in the upper panel, whereas an untransformed scale is used in the lower panel. The huge peak in the lower panel corresponds to the historic flood event in late December 2015.

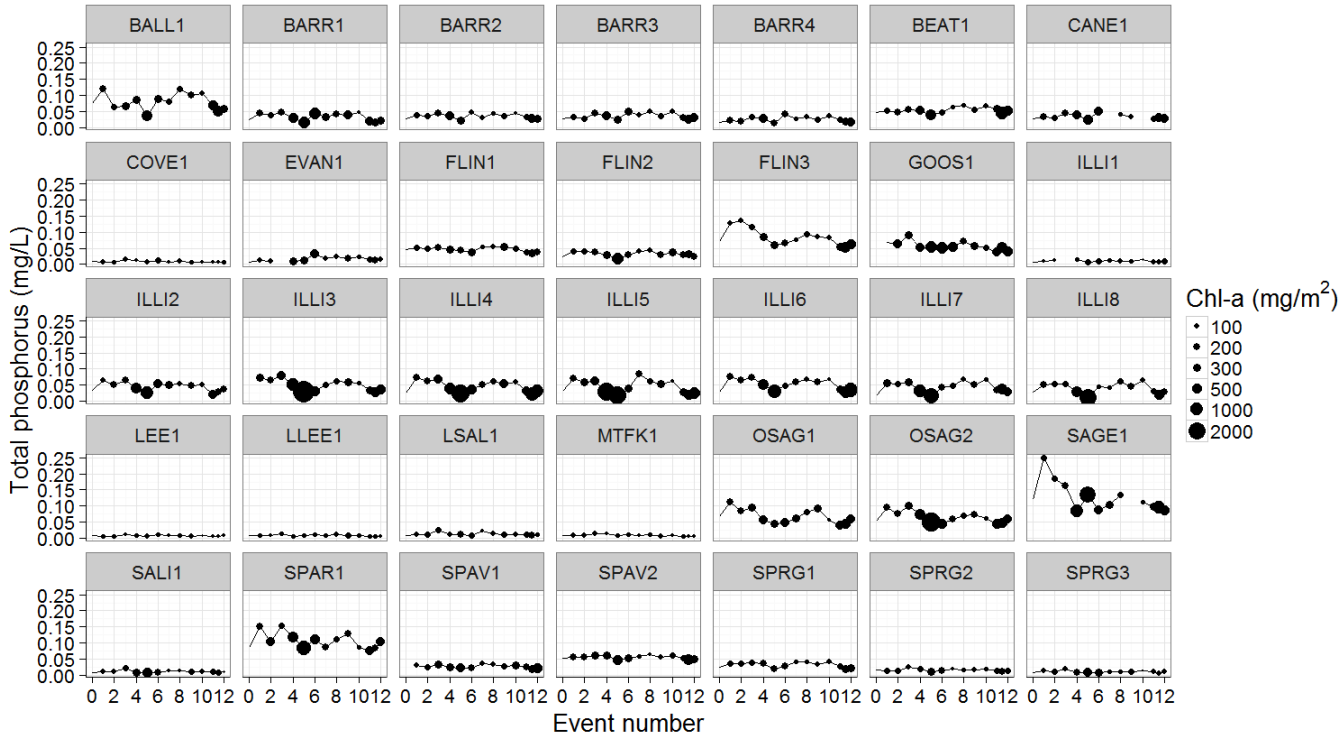

Figure 10. Temporal patterns of total phosphorus among the 35 study sites. Symbols are sized in relative proportion to benthic chlorophyll-a measured at the time of sampling. Note that, with the exception of SAGE1, which was effluent dominated, and to some degree, SPAR1 (also with a large proportion of base flow as wastewater effluent), most of the variability in TP over time within a site was related to whether there were high levels of benthic chlorophyll on the stream bottom at the time of sampling. In these cases, TP values declined sharply, very likely due to biological uptake. Sites with relatively low levels of TP and benthic chlorophyll-a throughout the study tended to have relatively consistent TP concentrations.

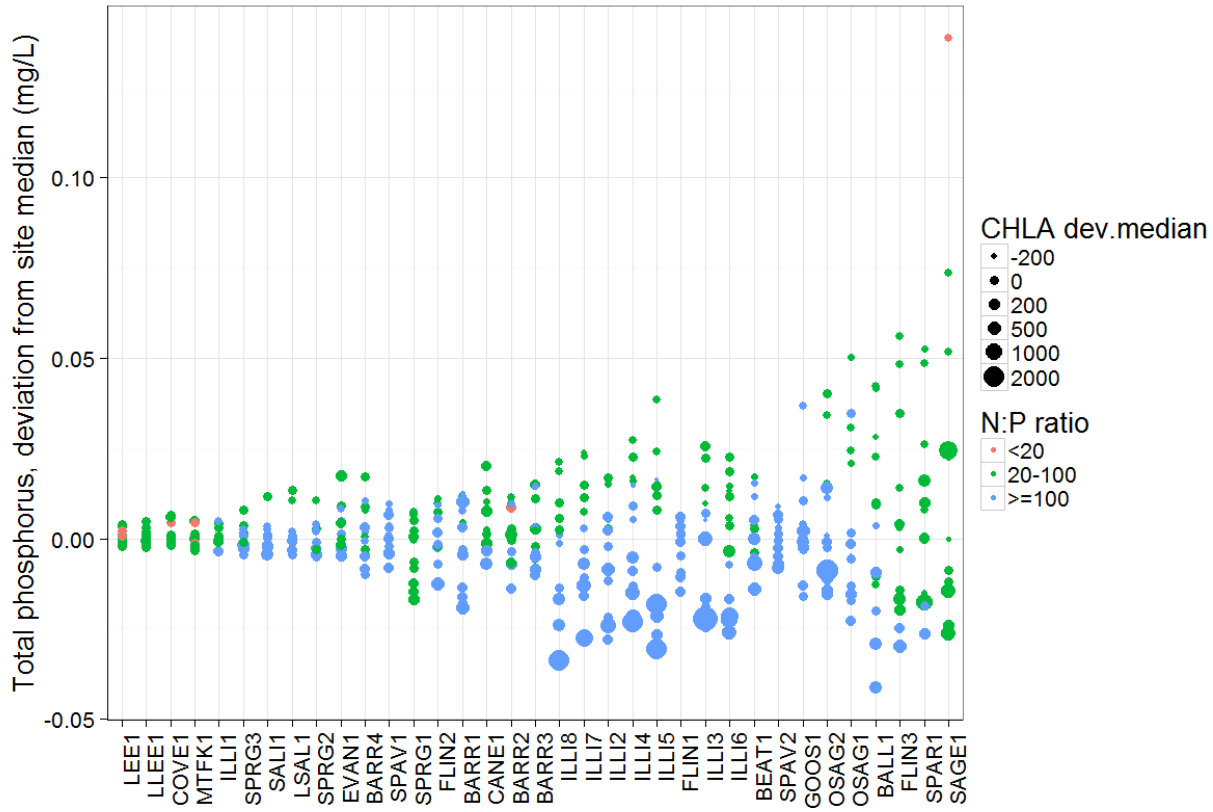

Figure 11. Dot plot of total phosphorus by sites (n=12 events), expressed as the deviation from the median 2-year concentration in mg/L. The 35 study sites are listed in rank order of their median 2-y TP concentrations. Each TP value is sized by the deviation from the site median for benthic chlorophyll-a; large values represent large, positive deviations from the typical level of chlorophyll at that site over the 2-year study. The colors represent the total nitrogen to total phosphorus ratio (N:P ratio) based on the measured TN and TP on that sampling event. N:P ratios <20 can be associated with N limiting conditions, whereas values above 20 increasingly demonstrate P limitation, or, at least, that there was a surplus of nitrogen relative to phosphorus. Note that in almost every case where benthic chlorophyll-a was much higher than the median (large dots), the total phosphorus value was lower, sometimes much lower, than the median. Further, under these conditions, the N:P ratio was >20 (green) and typically >100 (blue), but never <20 (orange). This implies that phosphorus, not nitrogen, was the driver of primary production among the study streams, although the high concentrations of nitrogen in these systems ensured that blooms were not restricted by N.

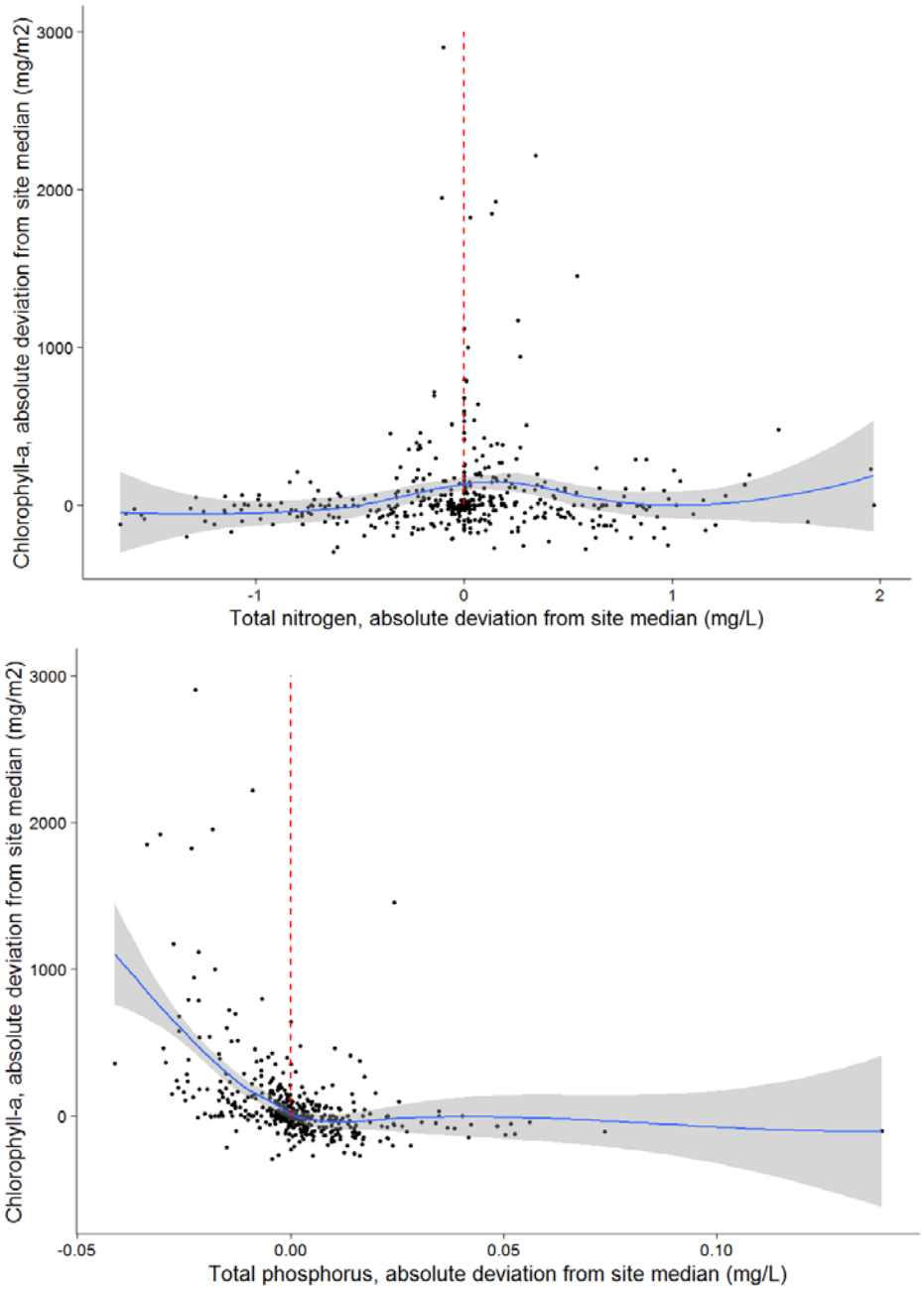

Figure 12. Plots of benthic chlorophyll-a in response to TN (upper) and TP (lower) deviations from site medians. The upper panel shows that the largest chlorophyll-a values were associated with mostly normal TN concentrations, with no relationship to benthic chlorophyll-a. The lower panel shows that almost all of the high chlorophyll-a levels corresponded to sharp reductions in TP. The fitted relationship shows that as TP levels were increasingly reduced, chlorophyll was at its highest. TP levels that are far above the median appear to be related to below normal levels of benthic chlorophyll-a.

Relationships between total phosphorus and algal biomass

Benthic chlorophyll-a varied markedly over time among the study sites (Figure 13). Levels of chlorophyll-a increased only slightly between June and October 2014, but increased dramatically during the months of December 2014 and February 2015 when a bloom of *Cladophora glomerata* was ongoing.

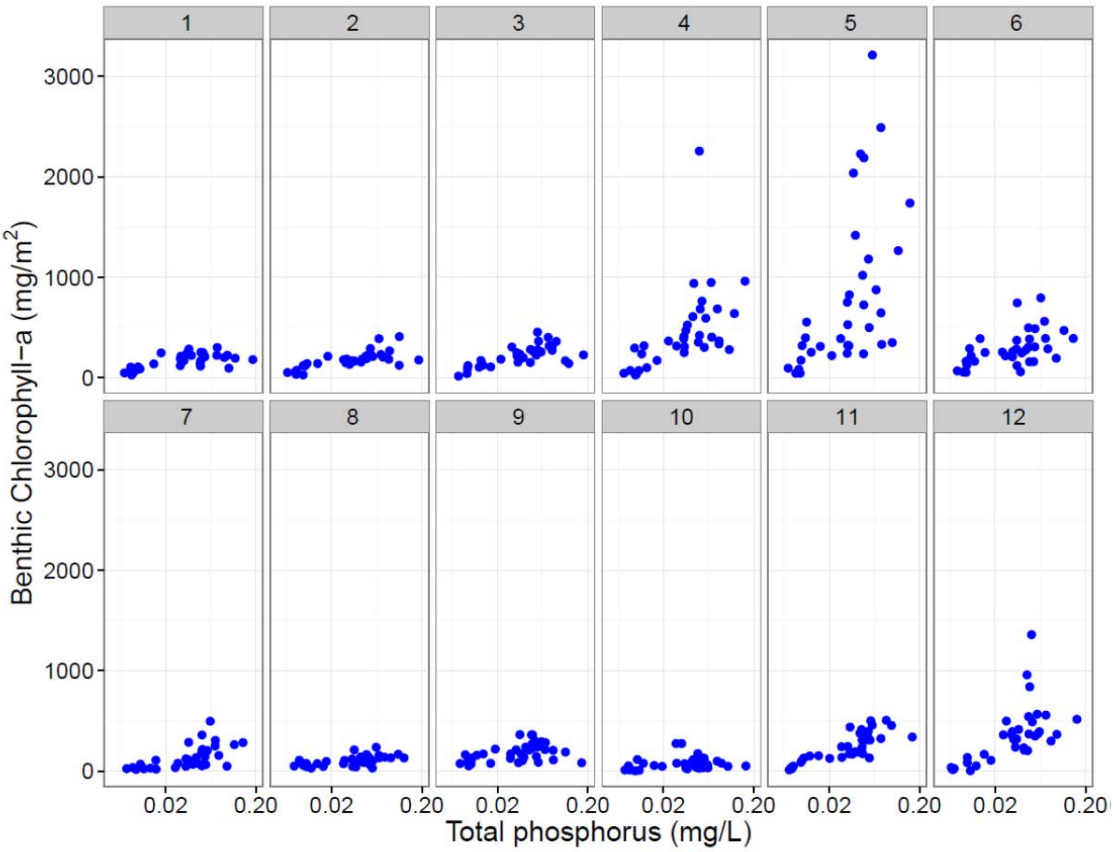

Figure 13. Relationship between benthic chlorophyll-a and 6-month mean TP across each event. Event 1 (June 2014) and 2 (August 2014) are based on 2 and 4 month mean TP values, respectively, because 6 month data was not available.

Benthic chlorophyll-a was reduced markedly by April 2015 following moderate storm flows that scoured much of the *Cladophora* off the stream bottom (Figure 13, 14). Reduction in benthic algal biomass continued through the summer and fall of 2015 (events 7-10). During this period, many large precipitation events resulted in very high stream flows and heavy scouring of algae, but often disproportionately among sites. Between event 10 (early December 2015) and 11 (early February 2016), the historic flood occurred that resulted in a complete scouring of substrate to the extent that channel morphology at most sites did not resemble previous conditions.

Despite the complete scouring following the historic flood, algal biomass recovered very quickly in by early February 2016, with some sites supporting levels up to 500 mg/m^2. However, filamentous green algae was not abundant during this event, and it appeared to be mostly dominated by diatoms and cyanobacteria. Further, due to a complete elimination of grazing macroinvertebrates, particularly pleurocerid snails, and the dormancy of the dominant vertebrate grazers (stonerollers, *Campostoma anomalum* and *Campostoma oligolepsis*; Taylor et al. 2012), the relationship between 6-month TP and algal biomass very closely resembled a theoretical growth-response curve, with a steep increase at low levels of TP and a gradual reduction in the slope (Figure 14, panel 11). By April 2016, *Cladophora glomerata* had become well established and contributed to even higher levels of algal biomass, with one site exceeding 1000 mg/m^2 (Figures 13 and 14, panel 12)

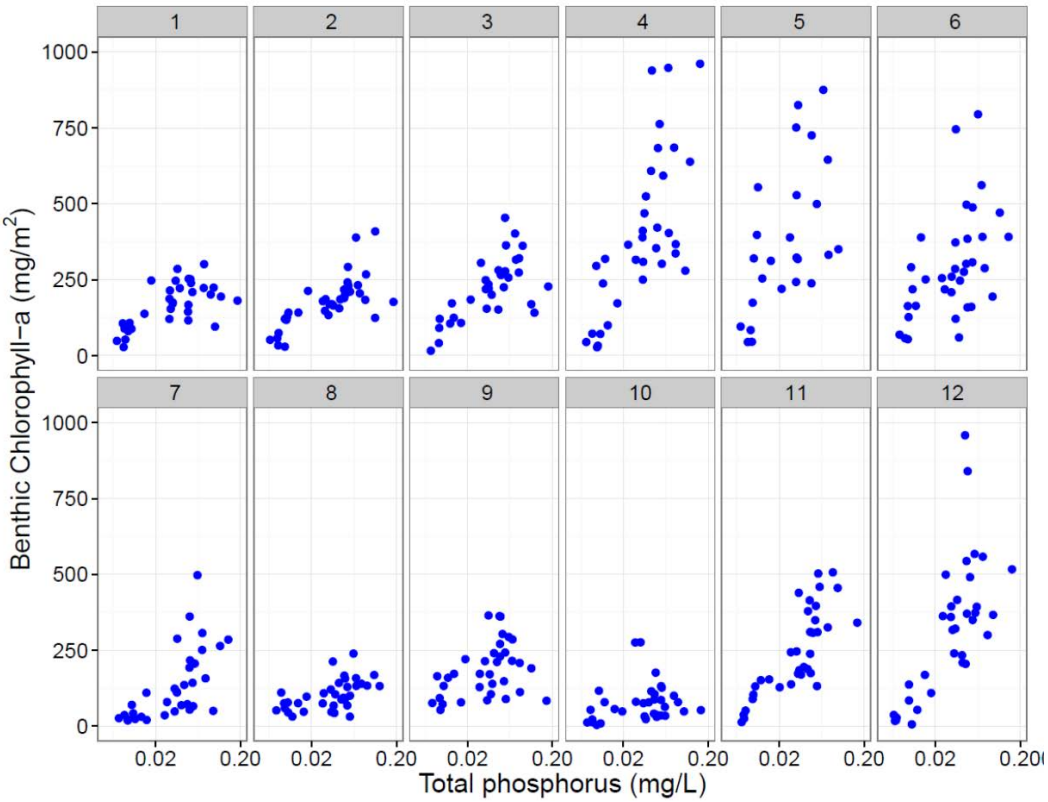

Figure 14. Relationship between benthic chlorophyll-a and 6-month mean TP across each event. This figure is identical to figure 13 except that the y-axis was truncated at 1000 mg/m^2 so that the relationship between TP and benthic chlorophyll-a during periods outside the massive *Cladophora* blooms could be better visualized.

Change point analysis: TP vs. benthic chlorophyll-a

The series of plots in the section *Temporal patterns in stream discharge, nutrients, and algal biomass* revealed the problem of relating nutrients to primary production or algal biomass. Despite the overall consistent levels of TP within a site over time, periods of high primary production depleted TP and caused the relationship between instantaneous measures of TP and algal biomass to break down. Thus, TP change points were estimated using means calculated at durations of 2, 4, 6, 8, 10, and 12 months. TP at these different durations were related to both instantaneous and mean chlorophyll-a (Figures 16 and 17, Tables 4 and 5).

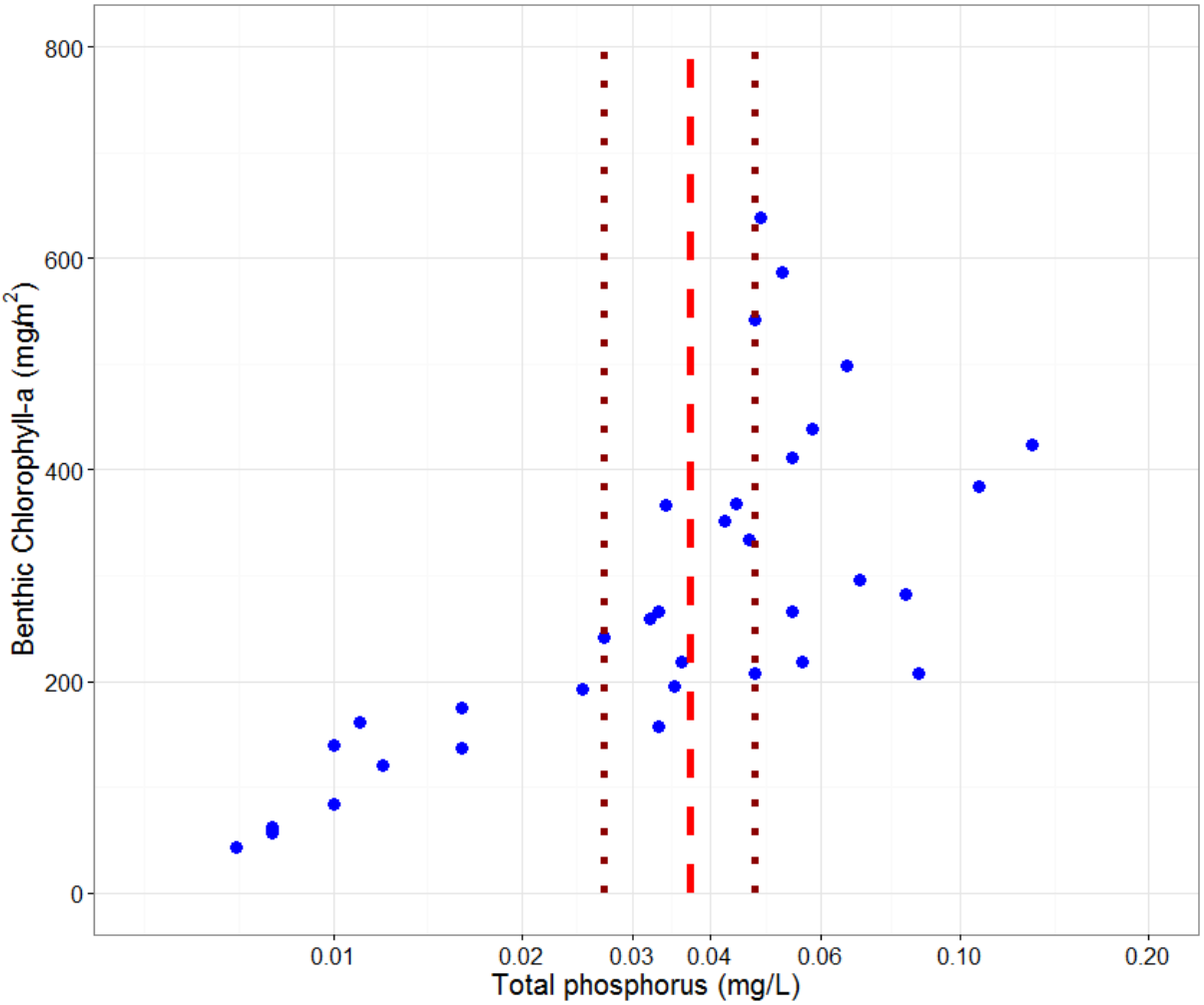

Figure 16. Two-year mean TP (April 2014-2016) vs. 2 year mean benthic chlorophyll-a. The dashed red line corresponds to 0.037 mg/L TP, whereas the dotted lines correspond to 0.027 and 0.047 mg/L TP, respectively.

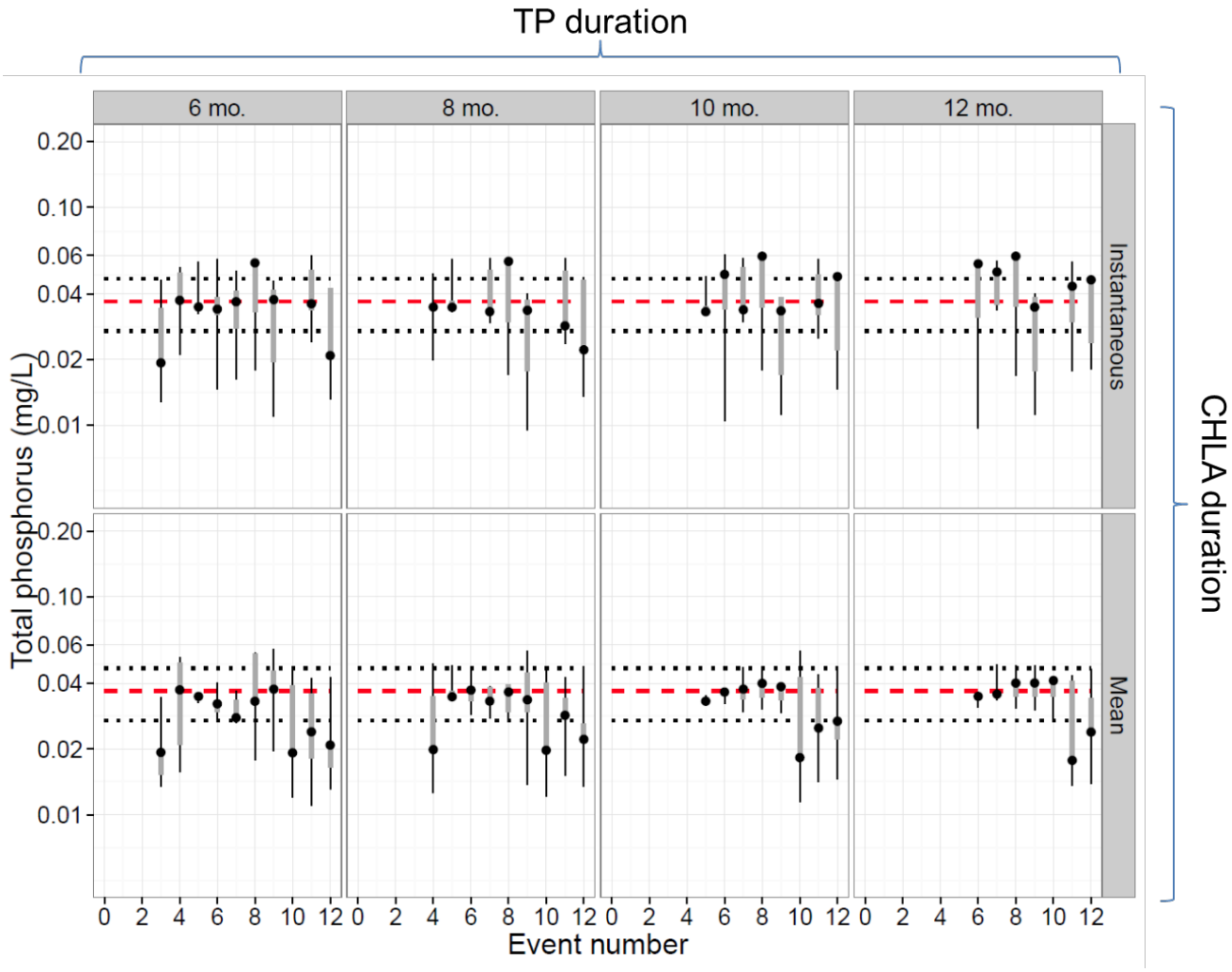

Figure 17. Total phosphorus change points in relation to benthic chlorophyll a. The columns represent 6, 8, 10, and 12 month TP durations, whereas the rows separate instantaneous and mean chlorophyll-a. Points correspond to the observed change point, gray bars span the 25-75% bootstrap quantiles, and black bars span the 5-95% bootstrap quantiles. The dashed red line is 0.037 mg/L TP, whereas the upper and lower dotted lines correspond to 0.027 and 0.047 mg/L. Results for 2 and 4 month TP are not shown, but are included in tables 4 and 5.

Table 4. Change points for 2, 4, 6, 8, 10, and 12 month mean total phosphorus in relation to instantaneous benthic chlorophyll-a.

				TP change points (mg/L)			Chlorophyll-a (mg/m2)		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	Chl-a Duration	Observed	Median (boot)	p-value	Mean (low)	Mean (high)	5% (boot)	25% (boot)	75% (boot)	95% (boot)
1	14-Jun	2 mo.	Instantaneous	0.016	0.016	0.001	81.9	200.8	0.010	0.013	0.016	0.029
2	14-Aug	2 mo.	Instantaneous	0.017	0.017	0.002	89.4	210.7	0.011	0.012	0.051	0.061
3	14-Oct	2 mo.	Instantaneous	0.022	0.022	0.001	97.1	260.7	0.016	0.017	0.027	0.055
4	14-Dec	2 mo.	Instantaneous	0.038	0.038	0.027	186.3	634.7	0.024	0.035	0.041	0.043
5	15-Feb	2 mo.	Instantaneous			0.071						
6	15-Apr	2 mo.	Instantaneous	0.029	0.029	0.023	194.4	383.9	0.017	0.028	0.029	0.044
7	15-Jun	2 mo.	Instantaneous	0.037	0.037	0.005	49.3	192.2	0.031	0.037	0.048	0.048
8	15-Aug	2 mo.	Instantaneous	0.061	0.055	0.031	87.7	143.0	0.019	0.031	0.061	0.069
9	15-Oct	2 mo.	Instantaneous	0.040	0.039	0.010	131.3	229.2	0.012	0.037	0.040	0.047
10	15-Dec	2 mo.	Instantaneous			0.275						
11	16-Feb	2 mo.	Instantaneous	0.031	0.031	0.001	111.7	318.8	0.018	0.027	0.042	0.044
12	16-Apr	2 mo.	Instantaneous	0.017	0.017	0.006	66.2	461.9	0.012	0.014	0.017	0.031
2	14-Aug	4 mo.	Instantaneous	0.016	0.016	0.001	89.4	210.7	0.010	0.011	0.043	0.051
3	14-Oct	4 mo.	Instantaneous	0.020	0.020	0.001	97.1	260.7	0.015	0.020	0.027	0.054
4	14-Dec	4 mo.	Instantaneous	0.037	0.037	0.020	213.1	659.4	0.022	0.037	0.038	0.049
5	15-Feb	4 mo.	Instantaneous	0.027	0.034	0.034	240.7	1105.2	0.025	0.027	0.035	0.065
6	15-Apr	4 mo.	Instantaneous	0.030	0.030	0.017	194.4	383.9	0.012	0.028	0.035	0.047
7	15-Jun	4 mo.	Instantaneous	0.030	0.033	0.008	49.3	192.2	0.028	0.030	0.041	0.053
8	15-Aug	4 mo.	Instantaneous	0.059	0.048	0.017	89.8	149.5	0.020	0.038	0.059	0.061
9	15-Oct	4 mo.	Instantaneous	0.037	0.037	0.019	128.7	221.6	0.012	0.029	0.040	0.046
10	15-Dec	4 mo.	Instantaneous			0.129						
11	16-Feb	4 mo.	Instantaneous	0.034	0.034	0.001	124.8	329.7	0.023	0.031	0.048	0.058
12	16-Apr	4 mo.	Instantaneous	0.021	0.021	0.008	66.2	464.7	0.013	0.016	0.038	0.042
3	14-Oct	6 mo.	Instantaneous	0.019	0.019	0.001	97.1	260.7	0.013	0.019	0.034	0.046
4	14-Dec	6 mo.	Instantaneous	0.037	0.037	0.010	213.1	659.4	0.021	0.037	0.050	0.053
5	15-Feb	6 mo.	Instantaneous	0.035	0.035	0.003	326.4	1319.5	0.033	0.033	0.035	0.056
6	15-Apr	6 mo.	Instantaneous	0.034	0.034	0.022	185.5	371.7	0.015	0.032	0.038	0.058
7	15-Jun	6 mo.	Instantaneous	0.037	0.034	0.010	74.1	208.4	0.016	0.028	0.041	0.051
8	15-Aug	6 mo.	Instantaneous	0.056	0.043	0.011	89.9	157.7	0.018	0.033	0.056	0.057
9	15-Oct	6 mo.	Instantaneous	0.038	0.038	0.025	130.3	220.5	0.011	0.020	0.042	0.046
10	15-Dec	6 mo.	Instantaneous			0.124						
11	16-Feb	6 mo.	Instantaneous	0.036	0.036	0.001	124.8	329.7	0.024	0.034	0.052	0.060
12	16-Apr	6 mo.	Instantaneous	0.021	0.021	0.003	66.2	461.9	0.013	0.021	0.043	0.043
4	14-Dec	8 mo.	Instantaneous	0.035	0.035	0.005	213.1	659.4	0.020	0.034	0.036	0.049
5	15-Feb	8 mo.	Instantaneous	0.035	0.035	0.007	297.1	1292.0	0.033	0.035	0.037	0.058
6	15-Apr	8 mo.	Instantaneous	0.014	0.036	0.055	142.7	336.1	0.010	0.014	0.048	0.059
7	15-Jun	8 mo.	Instantaneous	0.033	0.038	0.005	49.3	192.2	0.029	0.033	0.052	0.058
8	15-Aug	8 mo.	Instantaneous	0.056	0.055	0.012	89.9	157.7	0.017	0.030	0.056	0.056
9	15-Oct	8 mo.	Instantaneous	0.034	0.034	0.029	130.3	220.5	0.010	0.018	0.038	0.040
10	15-Dec	8 mo.	Instantaneous			0.106						
11	16-Feb	8 mo.	Instantaneous	0.029	0.043	0.001	93.1	310.1	0.024	0.029	0.051	0.058
12	16-Apr	8 mo.	Instantaneous	0.022	0.022	0.002	66.2	461.9	0.014	0.022	0.046	0.047
5	15-Feb	10 mo.	Instantaneous	0.033	0.033	0.004	297.1	1292.0	0.032	0.032	0.035	0.048
6	15-Apr	10 mo.	Instantaneous	0.049	0.041	0.025	229.9	410.6	0.010	0.034	0.049	0.061
7	15-Jun	10 mo.	Instantaneous	0.034	0.044	0.009	49.3	192.2	0.030	0.034	0.053	0.058
8	15-Aug	10 mo.	Instantaneous	0.060	0.045	0.011	89.9	157.7	0.018	0.035	0.060	0.060
9	15-Oct	10 mo.	Instantaneous	0.033	0.033	0.027	130.3	220.5	0.011	0.017	0.039	0.039
10	15-Dec	10 mo.	Instantaneous			0.116						
11	16-Feb	10 mo.	Instantaneous	0.036	0.040	0.001	120.5	322.8	0.025	0.032	0.049	0.058
12	16-Apr	10 mo.	Instantaneous	0.048	0.047	0.003	200.9	570.6	0.015	0.022	0.048	0.049
6	15-Apr	12 mo.	Instantaneous	0.055	0.035	0.042	245.9	447.2	0.010	0.031	0.054	0.055
7	15-Jun	12 mo.	Instantaneous	0.050	0.050	0.003	78.5	233.9	0.034	0.036	0.050	0.057
8	15-Aug	12 mo.	Instantaneous	0.060	0.048	0.014	89.9	157.7	0.017	0.035	0.060	0.060
9	15-Oct	12 mo.	Instantaneous	0.035	0.035	0.025	130.3	220.5	0.011	0.018	0.039	0.040
10	15-Dec	12 mo.	Instantaneous			0.215						
11	16-Feb	12 mo.	Instantaneous	0.043	0.041	0.001	150.7	350.1	0.018	0.030	0.043	0.056
12	16-Apr	12 mo.	Instantaneous	0.046	0.045	0.001	200.9	570.6	0.018	0.024	0.046	0.046

Table 5. Change points for 2, 4, 6, 8, 10, and 12 month mean total phosphorus in relation to mean benthic chlorophyll-a.

				TP change points (mg/L)			Chlorophyll-a (mg/m2)		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	Chl-a Duration	Observed (mg/L)	Median (boot)	p-value	Mean (low)	Mean (high)	5% (boot)	25% (boot)	75% (boot)	95% (boot)
2	14-Aug	2 mo.	Mean	0.017	0.017	0.001	85.7	209.5	0.011	0.013	0.017	0.057
3	14-Oct	2 mo.	Mean	0.022	0.022	0.001	91.0	235.7	0.017	0.017	0.022	0.055
4	14-Dec	2 mo.	Mean	0.038	0.038	0.012	164.9	451.9	0.019	0.035	0.041	0.043
5	15-Feb	2 mo.	Mean	0.016	0.016	0.047	182.5	811.9	0.010	0.016	0.021	0.032
6	15-Apr	2 mo.	Mean	0.026	0.026	0.017	235.8	736.2	0.022	0.024	0.027	0.027
7	15-Jun	2 mo.	Mean	0.037	0.037	0.017	121.9	278.8	0.015	0.035	0.040	0.048
8	15-Aug	2 mo.	Mean	0.053	0.053	0.017	76.6	165.2	0.025	0.036	0.061	0.061
9	15-Oct	2 mo.	Mean	0.040	0.040	0.004	101.2	180.0	0.019	0.039	0.040	0.047
10	15-Dec	2 mo.	Mean	0.019	0.019	0.009	76.1	150.1	0.007	0.015	0.033	0.042
11	16-Feb	2 mo.	Mean	0.023	0.023	0.001	65.1	199.2	0.011	0.018	0.023	0.031
12	16-Apr	2 mo.	Mean	0.017	0.017	0.001	77.4	382.5	0.012	0.017	0.019	0.031
2	14-Aug	4 mo.	Mean	0.016	0.016	0.001	85.7	209.5	0.010	0.012	0.016	0.051
3	14-Oct	4 mo.	Mean	0.020	0.020	0.001	87.3	226.1	0.015	0.016	0.020	0.038
4	14-Dec	4 mo.	Mean	0.037	0.037	0.004	152.9	379.5	0.016	0.022	0.038	0.049
5	15-Feb	4 mo.	Mean	0.029	0.034	0.017	203.0	667.5	0.020	0.027	0.035	0.035
6	15-Apr	4 mo.	Mean	0.027	0.027	0.010	218.3	696.2	0.020	0.026	0.028	0.028
7	15-Jun	4 mo.	Mean	0.030	0.030	0.006	173.7	574.8	0.027	0.029	0.030	0.037
8	15-Aug	4 mo.	Mean	0.038	0.038	0.011	105.7	224.5	0.015	0.035	0.045	0.059
9	15-Oct	4 mo.	Mean	0.037	0.037	0.003	82.7	177.9	0.019	0.037	0.046	0.060
10	15-Dec	4 mo.	Mean	0.019	0.019	0.003	72.1	138.7	0.011	0.019	0.042	0.048
11	16-Feb	4 mo.	Mean	0.023	0.023	0.001	84.2	202.0	0.011	0.018	0.031	0.039
12	16-Apr	4 mo.	Mean	0.021	0.021	0.001	65.4	288.1	0.013	0.016	0.021	0.038
3	14-Oct	6 mo.	Mean	0.019	0.019	0.001	87.3	226.1	0.013	0.015	0.019	0.034
4	14-Dec	6 mo.	Mean	0.037	0.037	0.001	160.4	385.1	0.016	0.021	0.050	0.053
5	15-Feb	6 mo.	Mean	0.035	0.035	0.003	245.4	759.2	0.033	0.033	0.035	0.035
6	15-Apr	6 mo.	Mean	0.032	0.032	0.006	219.4	718.2	0.027	0.029	0.032	0.040
7	15-Jun	6 mo.	Mean	0.028	0.031	0.004	157.2	548.5	0.026	0.028	0.034	0.037
8	15-Aug	6 mo.	Mean	0.033	0.037	0.008	105.7	224.5	0.018	0.033	0.055	0.056
9	15-Oct	6 mo.	Mean	0.038	0.038	0.002	84.3	176.9	0.020	0.038	0.046	0.058
10	15-Dec	6 mo.	Mean	0.019	0.030	0.004	72.1	138.7	0.012	0.019	0.039	0.047
11	16-Feb	6 mo.	Mean	0.024	0.024	0.001	84.2	202.0	0.011	0.018	0.036	0.042
12	16-Apr	6 mo.	Mean	0.021	0.021	0.001	65.4	288.2	0.013	0.016	0.021	0.043
4	14-Dec	8 mo.	Mean	0.020	0.020	0.001	101.7	315.4	0.013	0.020	0.035	0.049
5	15-Feb	8 mo.	Mean	0.035	0.035	0.002	201.5	615.8	0.033	0.034	0.037	0.048
6	15-Apr	8 mo.	Mean	0.037	0.036	0.002	243.6	657.2	0.029	0.033	0.037	0.046
7	15-Jun	8 mo.	Mean	0.033	0.033	0.002	176.8	589.6	0.028	0.033	0.038	0.039
8	15-Aug	8 mo.	Mean	0.037	0.037	0.002	172.0	473.4	0.028	0.030	0.040	0.040
9	15-Oct	8 mo.	Mean	0.034	0.034	0.010	111.8	223.1	0.014	0.030	0.045	0.056
10	15-Dec	8 mo.	Mean	0.020	0.033	0.003	62.1	145.5	0.012	0.020	0.040	0.047
11	16-Feb	8 mo.	Mean	0.029	0.029	0.001	82.3	183.8	0.015	0.029	0.034	0.043
12	16-Apr	8 mo.	Mean	0.022	0.022	0.001	79.7	267.8	0.014	0.022	0.026	0.046
5	15-Feb	10 mo.	Mean	0.033	0.033	0.001	185.1	537.6	0.032	0.032	0.033	0.035
6	15-Apr	10 mo.	Mean	0.037	0.037	0.001	203.0	567.2	0.032	0.036	0.037	0.037
7	15-Jun	10 mo.	Mean	0.038	0.037	0.001	211.7	564.9	0.030	0.034	0.038	0.047
8	15-Aug	10 mo.	Mean	0.040	0.039	0.003	194.7	534.5	0.031	0.035	0.040	0.048
9	15-Oct	10 mo.	Mean	0.039	0.038	0.001	173.8	444.1	0.029	0.033	0.039	0.039
10	15-Dec	10 mo.	Mean	0.018	0.030	0.005	80.7	185.8	0.011	0.018	0.043	0.056
11	16-Feb	10 mo.	Mean	0.025	0.032	0.001	72.0	178.5	0.014	0.025	0.038	0.044
12	16-Apr	10 mo.	Mean	0.027	0.027	0.001	84.5	240.8	0.015	0.022	0.027	0.048
6	15-Apr	12 mo.	Mean	0.035	0.035	0.001	189.1	510.7	0.031	0.035	0.035	0.035
7	15-Jun	12 mo.	Mean	0.036	0.036	0.001	170.9	497.6	0.034	0.036	0.036	0.049
8	15-Aug	12 mo.	Mean	0.040	0.040	0.003	188.8	492.0	0.031	0.035	0.040	0.048
9	15-Oct	12 mo.	Mean	0.040	0.039	0.002	185.9	483.2	0.030	0.035	0.040	0.048
10	15-Dec	12 mo.	Mean	0.041	0.038	0.002	158.9	383.4	0.028	0.035	0.041	0.041
11	16-Feb	12 mo.	Mean	0.018	0.030	0.001	81.3	204.3	0.014	0.018	0.041	0.043
12	16-Apr	12 mo.	Mean	0.024	0.024	0.001	71.0	226.7	0.014	0.024	0.034	0.046

Change point analysis: TP vs. Cladophora glomerata biovolume

Cladophora glomerata was the dominant filamentous green alga identified in the study. *Cladophora* is widely known as a nuisance species that proliferates with nutrient overenrichment (Dodds and Gudder 1992). Benthic algal biomass values that exceeded 200-300 mg/m^2 were typically associated with high levels of *Cladophora* biovolume.

Cladophora biovolume was very low to completely absent at relatively low levels of TP, but a clear, nonlinear change in its frequency and abundance occurred at moderate to high levels of TP (Figures 18 and 19, Table 6).

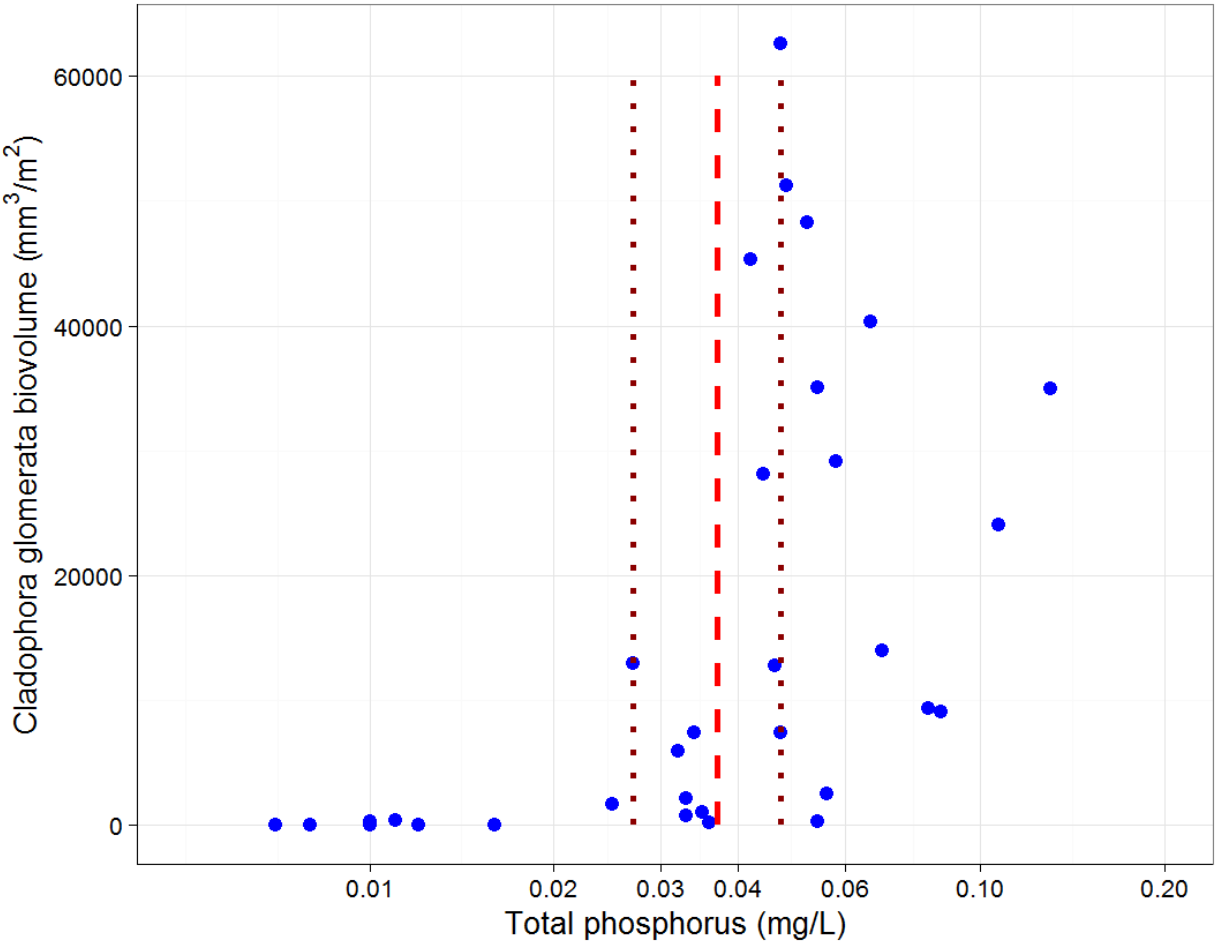

Figure 18. Two-year mean TP (April 2014-2016) vs. mean *Cladophora glomerata* biovolume from events 1, 3, 5, 6, 9, and 12. The dashed red line corresponds to 0.037 mg/L TP, whereas the dotted lines correspond to 0.027 and 0.047 mg/L TP, respectively.

Table 6. Change points for 6, 8, 10, and 12 month mean total phosphorus in relation to instantaneous and mean *Cladophora glomerata* biovolume. *Cladophora* biovolume was measured only on events 1, 3, 5, 6, 9, and 12.

				Change points (mg/L)			Biovolume (mm3/m2)		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	*Cladophora* Duration	Observed	Median (boot)	p-value	Mean (low)	Mean (high)	5% (boot)	25% (boot)	75% (boot)	95% (boot)
3	14-Oct	6 mo.	Instantaneous	0.048	0.048	0.001	406	9028	0.033	0.033	0.035	0.037
5	15-Feb	6 mo.	Instantaneous	0.037	0.035	0.004	3277	109728	0.033	0.034	0.037	0.041
6	15-Apr	6 mo.	Instantaneous			0.065						
9	15-Oct	6 mo.	Instantaneous	0.038	0.039	0.048	47	1031	0.035	0.037	0.052	0.066
12	16-Apr	6 mo.	Instantaneous	0.025	0.025	0.005	0	22428	0.018	0.024	0.042	0.043
5	15-Feb	8 mo.	Instantaneous	0.035	0.033	0.003	3277	109728	0.032	0.033	0.035	0.039
6	15-Apr	8 mo.	Instantaneous			0.058						
9	15-Oct	8 mo.	Instantaneous	0.034	0.034	0.035	47	1031	0.031	0.033	0.042	0.094
12	16-Apr	8 mo.	Instantaneous	0.026	0.026	0.032	0	22428	0.019	0.026	0.046	0.048
5	15-Feb	10 mo.	Instantaneous	0.035	0.035	0.007	3277	109728	0.033	0.033	0.035	0.048
6	15-Apr	10 mo.	Instantaneous									
9	15-Oct	10 mo.	Instantaneous	0.033	0.033	0.033	47	1031	0.030	0.033	0.039	0.085
12	16-Apr	10 mo.	Instantaneous	0.027	0.027	0.014	0	22428	0.019	0.026	0.048	0.049
6	15-Apr	12 mo.	Instantaneous			0.074						
9	15-Oct	12 mo.	Instantaneous	0.035	0.035	0.029	47	1031	0.032	0.034	0.047	0.083
12	16-Apr	12 mo.	Instantaneous	0.024	0.024	0.033	0	22428	0.015	0.021	0.046	0.048
3	14-Oct	6 mo.	Mean	0.048	0.036	0.007	1515	6848	0.016	0.033	0.050	0.055
5	15-Feb	6 mo.	Mean	0.035	0.034	0.002	1771	58932	0.033	0.034	0.035	0.040
6	15-Apr	6 mo.	Mean	0.032	0.031	0.014	1235	49174	0.027	0.031	0.032	0.038
9	15-Oct	6 mo.	Mean	0.051	0.051	0.017	492	6327	0.049	0.050	0.052	0.058
12	16-Apr	6 mo.	Mean	0.043	0.039	0.042	2229	15714	0.018	0.024	0.043	0.046
5	15-Oct	8 mo.	Mean	0.040	0.039	0.004	1334	40199	0.035	0.039	0.040	0.042
6	16-Apr	8 mo.	Mean	0.046	0.046	0.033	2229	15714	0.019	0.026	0.046	0.049
9	15-Feb	8 mo.	Mean	0.037	0.035	0.003	1980	40855	0.033	0.034	0.037	0.041
12	15-Apr	8 mo.	Mean	0.037	0.037	0.001	1343	42562	0.036	0.037	0.037	0.039
5	15-Feb	10 mo.	Mean	0.035	0.033	0.001	1980	40855	0.032	0.033	0.035	0.039
6	15-Apr	10 mo.	Mean	0.038	0.037	0.001	1605	33097	0.036	0.036	0.038	0.048
9	15-Oct	10 mo.	Mean	0.039	0.038	0.001	1334	40199	0.036	0.038	0.039	0.041
12	16-Apr	10 mo.	Mean	0.048	0.047	0.015	2248	16650	0.019	0.027	0.048	0.049
6	15-Apr	12 mo.	Mean	0.037	0.035	0.001	1605	33097	0.034	0.035	0.036	0.037
9	15-Oct	12 mo.	Mean	0.039	0.038	0.001	1069	32183	0.036	0.038	0.039	0.041
12	16-Apr	12 mo.	Mean	0.046	0.046	0.025	1907	14665	0.021	0.040	0.046	0.049
12	16-Apr	24 mo.	Mean	0.039	0.039	0.002	1832	26752	0.035	0.035	0.039	0.047

Figure 19. Photograph of *Cladophora glomerata* covering the stream bottom of the Illinois River at Tahlequah (ILLI8), February 2015.

Change point analysis: TP vs. nuisance taxa proportion of total biovolume

Five genera of filamentous green algae that occurred in our data set were classified as nuisance taxa: *Cladophora, Oedogonium, Rhizoclonium, Spirogyra, and Hydrodictyon.* Although *Cladophora* represented most of the total nuisance biovolume (>95%), there were a few sites that had blooms of other taxa during the 2 year study. The committee recommended that the analysis be conducted on the proportion of the total biovolume as nuisance taxa as a complementary but different way of examining the data (Figure 20, Table 7). Because diatoms were identified on only 4 events compared to 6 events for soft algae, proportions were calculated based on the total soft-algae biovolume. A binomial form of change point analysis was used for these data, which is appropriate for proportion data (Zuur et al. 2009).

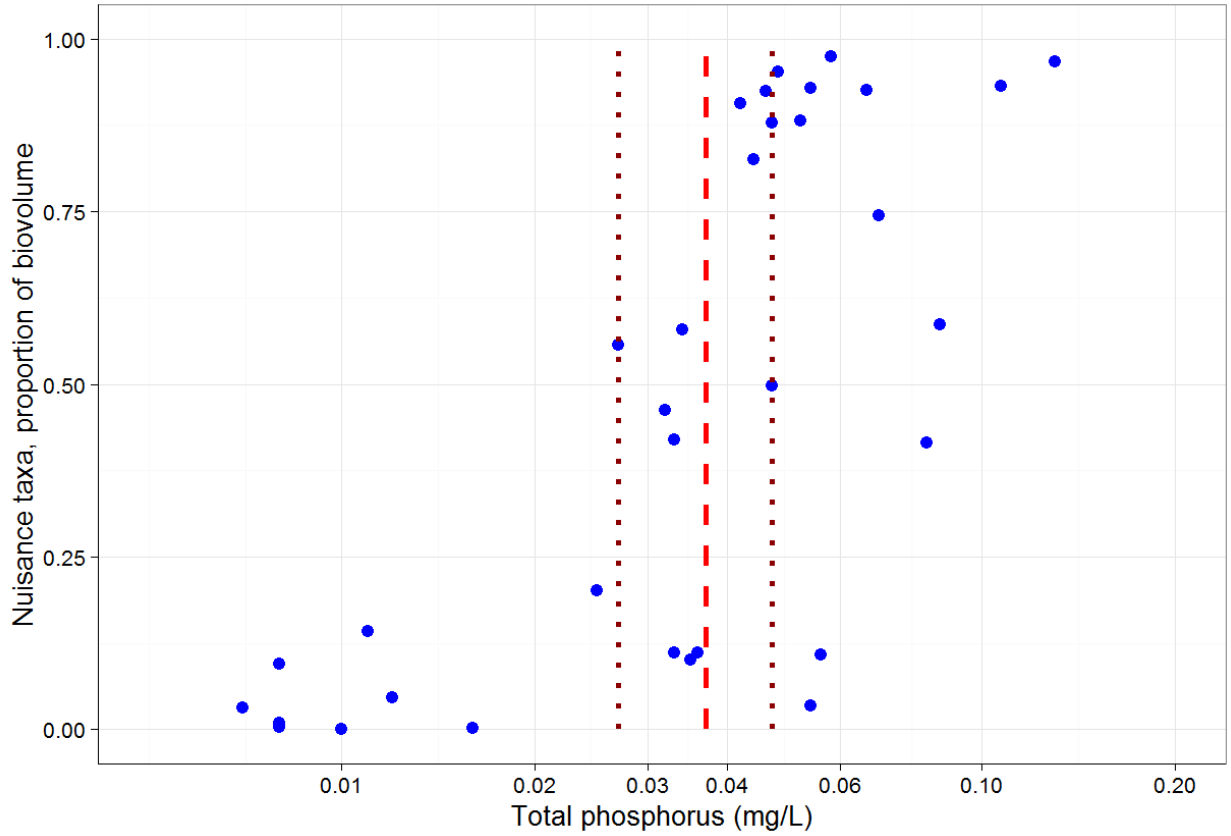

Figure 20. Two-year mean TP (April 2014-2016) vs. mean nuisance taxa proportion from events 1, 3, 5, 6, 9, and 12. The dashed red line corresponds to 0.037 mg/L TP, whereas the dotted lines correspond to 0.027 and 0.047 mg/L TP, respectively.

Table 7. Change points for 6, 8, 10, and 12 month mean total phosphorus in relation to instantaneous and mean nuisance taxa proportion of total biovolume. Soft algal species composition was measured only on events 1, 3, 5, 6, 9, and 12.

				TP change points (mg/L)			Nuisance proportion		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	Nuisance Duration	Observed	Median (boot)	p-value	Mean (low)	Mean (high)	5% (boot)	25% (boot)	75% (boot)	95% (boot)
	14-Oct	6 mo.	Instantaneous	0.074	0.073	0.001	0.238	0.781	0.047	0.052	0.077	0.085
	15-Feb	6 mo.	Instantaneous	0.035	0.035	0.005	0.153	0.856	0.033	0.034	0.052	0.056
	15-Apr	6 mo.	Instantaneous			ns						
	15-Oct	6 mo.	Instantaneous	0.051	0.051	0.036	0.112	0.330	0.045	0.051	0.052	0.097
	16-Apr	6 mo.	Instantaneous	0.043	0.042	0.002	0.148	0.878	0.035	0.038	0.043	0.058
	15-Feb	8 mo.	Instantaneous	0.035	0.036	0.002	0.106	0.861	0.033	0.034	0.056	0.058
	15-Apr	8 mo.	Instantaneous			ns						
	15-Oct	8 mo.	Instantaneous			ns						
	16-Apr	8 mo.	Instantaneous	0.046	0.046	0.005	0.148	0.878	0.038	0.042	0.047	0.060
	15-Feb	10 mo.	Instantaneous	0.033	0.033	0.004	0.106	0.861	0.032	0.033	0.048	0.058
	15-Apr	10 mo.	Instantaneous			ns						
	15-Oct	10 mo.	Instantaneous			ns						
	16-Apr	10 mo.	Instantaneous	0.047	0.047	0.004	0.148	0.878	0.036	0.043	0.048	0.060
	15-Apr	12 mo.	Instantaneous			ns						
	15-Oct	12 mo.	Instantaneous			ns						
	16-Apr	12 mo.	Instantaneous	0.046	0.046	0.003	0.148	0.878	0.037	0.043	0.046	0.058
	14-Oct	6 mo.	Mean	0.052	0.057	0.049	0.116	0.464	0.036	0.051	0.074	0.113
	15-Feb	6 mo.	Mean	0.035	0.035	0.004	0.154	0.840	0.033	0.034	0.048	0.056
	15-Apr	6 mo.	Mean	0.040	0.040	0.040	0.273	0.800	0.028	0.037	0.055	0.064
	15-Oct	6 mo.	Mean	0.058	0.057	0.096	0.263	0.579	0.036	0.050	0.059	0.102
	16-Apr	6 mo.	Mean	0.043	0.042	0.015	0.223	0.777	0.031	0.038	0.044	0.058
	15-Feb	8 mo.	Mean	0.035	0.049	0.011	0.102	0.701	0.033	0.035	0.056	0.105
	15-Apr	8 mo.	Mean	0.037	0.037	0.003	0.160	0.820	0.036	0.037	0.048	0.059
	15-Oct	8 mo.	Mean			ns						
	16-Apr	8 mo.	Mean	0.046	0.046	0.010	0.223	0.777	0.037	0.040	0.047	0.061
	15-Feb	10 mo.	Mean	0.033	0.035	0.009	0.102	0.701	0.032	0.033	0.054	0.100
	15-Apr	10 mo.	Mean	0.037	0.038	0.002	0.120	0.821	0.036	0.036	0.048	0.055
	15-Oct	10 mo.	Mean			ns						
	16-Apr	10 mo.	Mean	0.048	0.048	0.005	0.238	0.794	0.036	0.046	0.048	0.060
	15-Apr	12 mo.	Mean	0.035	0.036	0.009	0.107	0.703	0.035	0.035	0.055	0.095
	15-Oct	12 mo.	Mean	0.039	0.039	0.003	0.170	0.803	0.036	0.038	0.040	0.056
	16-Apr	12 mo.	Mean	0.046	0.046	0.020	0.274	0.754	0.037	0.042	0.047	0.059
	16-Apr	24 mo.	Mean	0.039	0.039	0.005	0.179	0.734	0.035	0.036	0.040	0.061

TITAN: TP vs. algal community composition

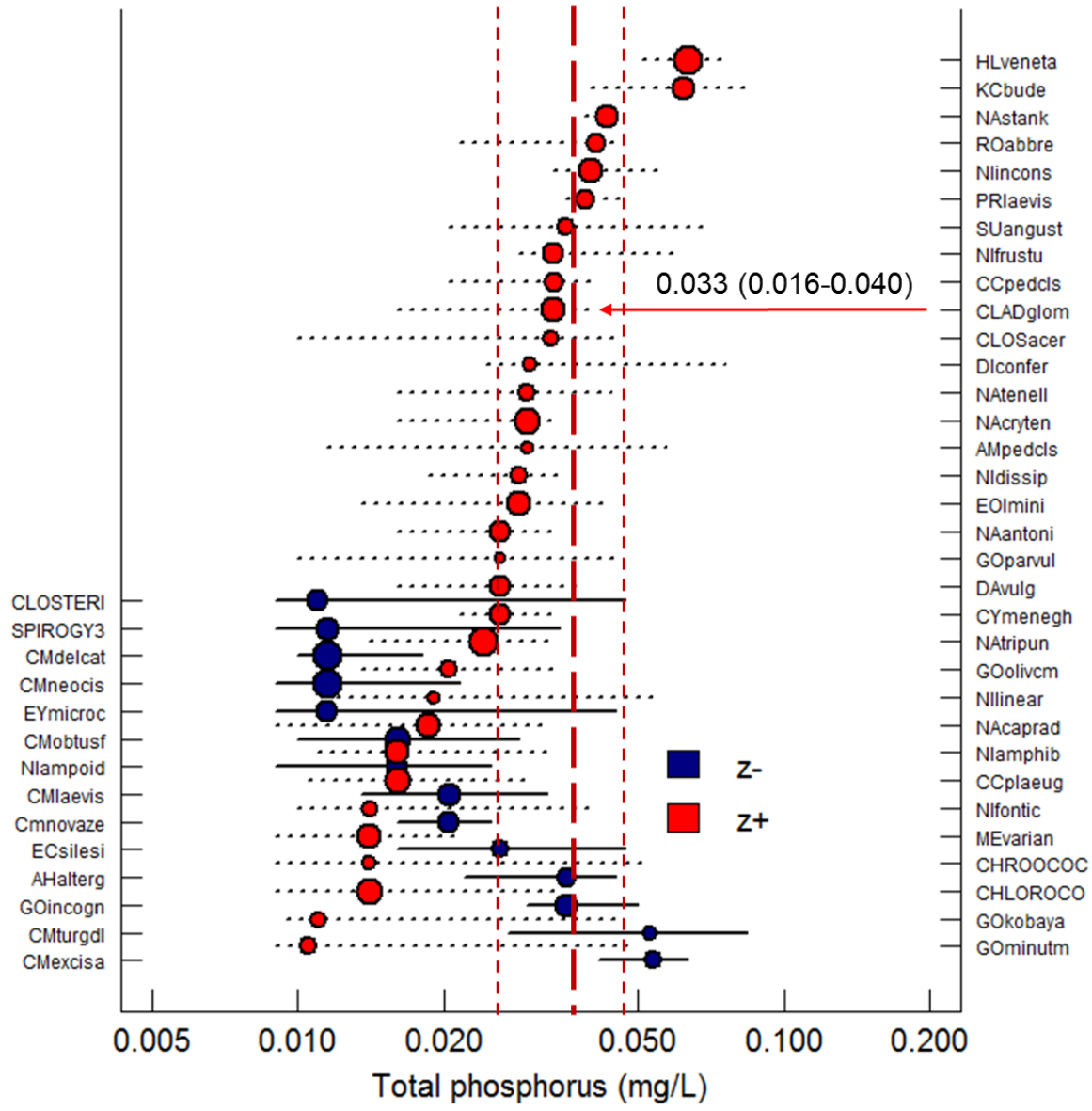

Figure 21. Results of TITAN using two-year mean TP (April 2014-2016) as the predictor vs. mean biovolume of all taxa that occurred at least 3 times from events 1, 3, 5, 6, 9, and 12. Shown are pure and reliable threshold indicator taxa. Negative responding taxa are listed on the left y-axis and marked by dark blue points, whereas positive responding taxa are on the right y-axis and marked by red points. Points are located at change point. Error bars represent the 5-95% bootstrap quantile intervals. The community-level threshold for negative-responding taxa (sumz-) was 0.021 (0.010-0.025) mg/L, whereas the positive-responding community threshold was also 0.021 mg/L, but had higher bootstrap quantile intervals (0.016-0.033). The vertical, red dashed line corresponds to 0.037 mg/L TP, whereas the vertical dotted lines correspond to 0.027 and 0.047 mg/L TP, respectively.

Table 8. TITAN community-level negative (declining taxa only) change points for 6, 8, 10, and 12 month mean total phosphorus in relation to instantaneous and mean taxa biovolumes.

					TP change points (mg/L)		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	Taxa Duration	Response direction	Observed	Median(boot)	5% (boot)	10% (boot)	90% (boot)	95% (boot)
1	14-Jun	6 mo.	Instantaneous	sumz- (negative)	0.019	0.019	0.016	0.017	0.025	0.025
3	14-Oct	6 mo.	Instantaneous	sumz- (negative)	0.013	0.013	0.010	0.011	0.035	0.037
5	15-Feb	6 mo.	Instantaneous	sumz- (negative)	0.025	0.025	0.013	0.013	0.029	0.033
6	15-Apr	6 mo.	Instantaneous	sumz- (negative)	0.017	0.016	0.012	0.012	0.021	0.022
9	15-Oct	6 mo.	Instantaneous	sumz- (negative)	0.026	0.026	0.016	0.021	0.029	0.030
12	16-Apr	6 mo.	Instantaneous	sumz- (negative)	0.010	0.013	0.010	0.010	0.023	0.027
5	15-Feb	8 mo.	Instantaneous	sumz- (negative)	0.033	0.033	0.013	0.013	0.035	0.037
6	15-Apr	8 mo.	Instantaneous	sumz- (negative)	0.012	0.014	0.011	0.011	0.020	0.021
9	15-Oct	8 mo.	Instantaneous	sumz- (negative)	0.024	0.024	0.014	0.015	0.027	0.029
12	16-Apr	8 mo.	Instantaneous	sumz- (negative)	0.011	0.013	0.011	0.011	0.029	0.037
5	15-Feb	10 mo.	Instantaneous	sumz- (negative)	0.024	0.024	0.012	0.012	0.032	0.033
6	15-Apr	10 mo.	Instantaneous	sumz- (negative)	0.012	0.013	0.010	0.011	0.020	0.021
9	16-Apr	10 mo.	Instantaneous	sumz- (negative)	0.023	0.023	0.014	0.015	0.027	0.029
12	15-Oct	10 mo.	Instantaneous	sumz- (negative)	0.011	0.012	0.009	0.010	0.033	0.034
6	15-Apr	12 mo.	Instantaneous	sumz- (negative)	0.011	0.012	0.011	0.011	0.020	0.021
9	15-Oct	12 mo.	Instantaneous	sumz- (negative)	0.023	0.023	0.015	0.017	0.028	0.030
12	16-Apr	12 mo.	Instantaneous	sumz- (negative)	0.012	0.012	0.009	0.010	0.030	0.035
3	14-Oct	6 mo.	Mean	sumz- (negative)	0.019	0.019	0.010	0.011	0.025	0.025
5	15-Feb	6 mo.	Mean	sumz- (negative)	0.025	0.028	0.013	0.021	0.037	0.037
6	15-Apr	6 mo.	Mean	sumz- (negative)	0.020	0.020	0.010	0.013	0.024	0.025
9	15-Oct	6 mo.	Mean	sumz- (negative)	0.020	0.020	0.014	0.016	0.027	0.029
12	16-Apr	6 mo.	Mean	sumz- (negative)	0.021	0.021	0.010	0.011	0.024	0.027
5	15-Feb	8 mo.	Mean	sumz- (negative)	0.025	0.028	0.013	0.021	0.037	0.039
6	15-Apr	8 mo.	Mean	sumz- (negative)	0.016	0.016	0.011	0.012	0.021	0.023
9	15-Oct	8 mo.	Mean	sumz- (negative)	0.024	0.021	0.012	0.013	0.024	0.027
12	16-Apr	8 mo.	Mean	sumz- (negative)	0.022	0.022	0.011	0.013	0.029	0.029
5	15-Feb	10 mo.	Mean	sumz- (negative)	0.019	0.019	0.011	0.016	0.024	0.025
6	15-Apr	10 mo.	Mean	sumz- (negative)	0.014	0.015	0.010	0.012	0.026	0.034
9	15-Oct	10 mo.	Mean	sumz- (negative)	0.023	0.021	0.010	0.011	0.023	0.026
12	16-Apr	10 mo.	Mean	sumz- (negative)	0.011	0.022	0.011	0.011	0.025	0.027
6	15-Apr	12 mo.	Mean	sumz- (negative)	0.020	0.019	0.012	0.012	0.021	0.023
9	15-Oct	12 mo.	Mean	sumz- (negative)	0.023	0.018	0.013	0.013	0.023	0.025
12	16-Apr	12 mo.	Mean	sumz- (negative)	0.018	0.018	0.011	0.012	0.024	0.024
12	16-Apr	24 mo.	Mean	sumz- (negative)	0.021	0.021	0.010	0.011	0.024	0.025

Table 9. TITAN community-level positive (increasing taxa only) change points for 6, 8, 10, and 12 month mean total phosphorus in relation to instantaneous and mean taxa biovolumes.

					TP change points (mg/L)		Bootstrap quantiles (mg/L)			
Event	Date	TP Duration	Taxa Duration	Response direction	Observed	Median(boot)	5% (boot)	10% (boot)	90% (boot)	95% (boot)
1	14-Jun	6 mo.	Instantaneous	sumz+ (positive)	0.035	0.034	0.018	0.019	0.056	0.056
3	14-Oct	6 mo.	Instantaneous	sumz+ (positive)	0.056	0.056	0.042	0.051	0.064	0.070
5	15-Feb	6 mo.	Instantaneous	sumz+ (positive)	0.020	0.025	0.019	0.020	0.035	0.037
6	15-Apr	6 mo.	Instantaneous	sumz+ (positive)	0.038	0.034	0.022	0.027	0.042	0.042
9	15-Oct	6 mo.	Instantaneous	sumz+ (positive)	0.050	0.038	0.020	0.021	0.050	0.050
12	16-Apr	6 mo.	Instantaneous	sumz+ (positive)	0.025	0.021	0.018	0.020	0.043	0.043
5	15-Feb	8 mo.	Instantaneous	sumz+ (positive)	0.025	0.025	0.019	0.019	0.037	0.039
6	15-Apr	8 mo.	Instantaneous	sumz+ (positive)	0.039	0.038	0.027	0.029	0.043	0.048
9	15-Oct	8 mo.	Instantaneous	sumz+ (positive)	0.034	0.034	0.016	0.018	0.043	0.044
12	16-Apr	8 mo.	Instantaneous	sumz+ (positive)	0.022	0.026	0.019	0.021	0.046	0.047
5	15-Feb	10 mo.	Instantaneous	sumz+ (positive)	0.024	0.024	0.019	0.019	0.035	0.036
6	15-Apr	10 mo.	Instantaneous	sumz+ (positive)	0.037	0.037	0.031	0.032	0.049	0.049
9	16-Apr	10 mo.	Instantaneous	sumz+ (positive)	0.039	0.038	0.017	0.021	0.042	0.042
12	15-Oct	10 mo.	Instantaneous	sumz+ (positive)	0.022	0.027	0.019	0.021	0.047	0.048
6	15-Apr	12 mo.	Instantaneous	sumz+ (positive)	0.035	0.035	0.029	0.031	0.041	0.043
9	15-Oct	12 mo.	Instantaneous	sumz+ (positive)	0.040	0.035	0.015	0.018	0.048	0.048
12	16-Apr	12 mo.	Instantaneous	sumz+ (positive)	0.024	0.024	0.020	0.021	0.046	0.046
3	14-Oct	6 mo.	Mean	sumz+ (positive)	0.019	0.021	0.017	0.018	0.035	0.040
5	15-Feb	6 mo.	Mean	sumz+ (positive)	0.037	0.037	0.023	0.025	0.049	0.051
6	15-Apr	6 mo.	Mean	sumz+ (positive)	0.029	0.035	0.023	0.025	0.043	0.048
9	15-Oct	6 mo.	Mean	sumz+ (positive)	0.049	0.047	0.032	0.033	0.050	0.050
12	16-Apr	6 mo.	Mean	sumz+ (positive)	0.016	0.021	0.014	0.016	0.031	0.035
5	15-Feb	8 mo.	Mean	sumz+ (positive)	0.033	0.035	0.021	0.024	0.049	0.051
6	15-Apr	8 mo.	Mean	sumz+ (positive)	0.039	0.037	0.020	0.026	0.043	0.046
9	15-Oct	8 mo.	Mean	sumz+ (positive)	0.030	0.030	0.021	0.024	0.042	0.043
12	16-Apr	8 mo.	Mean	sumz+ (positive)	0.022	0.024	0.015	0.017	0.035	0.041
5	15-Feb	10 mo.	Mean	sumz+ (positive)	0.019	0.021	0.016	0.017	0.032	0.033
6	15-Apr	10 mo.	Mean	sumz+ (positive)	0.037	0.037	0.024	0.026	0.041	0.049
9	15-Oct	10 mo.	Mean	sumz+ (positive)	0.029	0.029	0.022	0.023	0.040	0.041
12	16-Apr	10 mo.	Mean	sumz+ (positive)	0.027	0.023	0.016	0.017	0.036	0.041
6	15-Apr	12 mo.	Mean	sumz+ (positive)	0.025	0.025	0.019	0.020	0.035	0.035
9	15-Oct	12 mo.	Mean	sumz+ (positive)	0.028	0.028	0.018	0.018	0.040	0.042
12	16-Apr	12 mo.	Mean	sumz+ (positive)	0.024	0.024	0.016	0.018	0.042	0.042
12	16-Apr	24 mo.	Mean	sumz+ (positive)	0.021	0.024	0.016	0.019	0.029	0.033

Reference value threshold approach

We related biovolume of *Cladophora glomerata*, which was measured during events 1, 3, 5, 6, 9, and 12, to benthic chlorophyll-a, which was measured during all 12 events, to evaluate whether there was a level of benthic chlorophyll that corresponded to a nonlinear increase in *Cladophora*. The rationale was that (1) we did not have biovolume of *Cladophora* for all events, because this requires manual microscopic estimation by an expert taxonomist, a tedious and expensive process beyond the budget of this study, (2) "nuisance" levels defined by the literature are subjective and context dependent, and (3) some of our sites with low phosphorus consistently yielded benthic chlorophyll-a levels that approached or exceeded literature values for "nuisance" conditions (>150-200 mg/m^2), yet virtually none of this algal biomass was *Cladophora* or other nuisance species of filamentous green algae, and (4) our sampling protocol required large substrates (10-20 cm) for chlorophyll-a estimation, whereas most other protocols do not specify substrate size and thus are more likely to include smaller substrates that are much more prone to tumbling and scouring and thus would bias chlorophyll-a estimates downward, especially at sites dominated by small gravel.

Graphical visualization of the relationship between mean *Cladophora* biovolume and benthic chlorophyll-a suggested that segmented regression would be the most appropriate method for estimating the level of algal biomass that corresponded to a shift to the dominant nuisance species in the Designated Scenic Rivers. This particular method is generally not appropriate for most types of ecological data because it requires that the relationship between two variables can be represented by two or more linear segments that conform to parametric assumptions of normality and homoscedasticity. However, in this instance, these two variables were dependent on each other and exhibited a relationship that was ideal for segmented regression. Lack of independence was not an issue here because we were not testing a hypothesis that required this assumption.

Results of these analyses (Figure 22) indicated that 290 and 183 mg/m2 benthic chlorophyll-a were levels corresponding to a shift from essentially no *Cladophora* to a linear increase in *Cladophora* biovolume during years 1 and 2, respectively. Year 1 was not dry, but lacked significant scouring events, particularly during fall 2014 when the *Cladophora* bloom began to take hold. Year 2 was wet and had many significant scouring events including the historic flood in December 2015.

Based on these results, after rounding up/down to account for statistical uncertainty (see confidence limits, Figure 22), we agreed that 150-200 mg/m^2 likely represented the lower end of potential nuisance levels of algal biomass in the Designated Scenic Rivers during a wet year, whereas levels above 300 mg/m^2 should be considered nuisance levels under most conditions, acknowledging that a few sites with the lowest levels of TP in the region achieved benthic chlorophyll-a >300 mg/m^2 in February 2015, an event marking the end of several months of relatively stable flow.

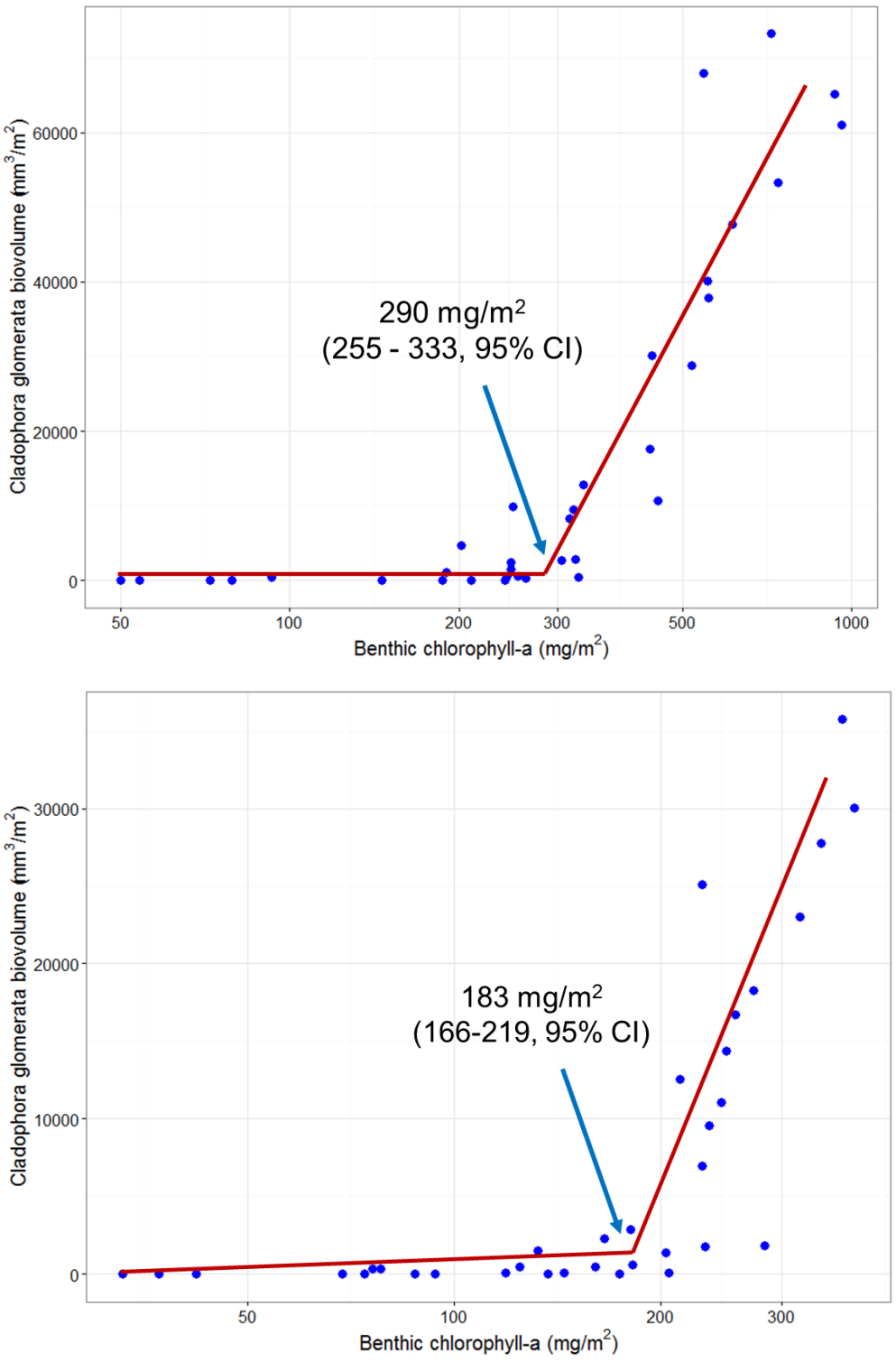

Figure 22. Results of segmented regression relating mean *Cladophora glomerata* biovolume to mean benthic chlorophyll-a during year 1 (upper panel) and year 2 (lower panel). Year 1 data represents a year with very stable flows overall, whereas year 2 represents a wet year with many scouring flows, including an historic flood.

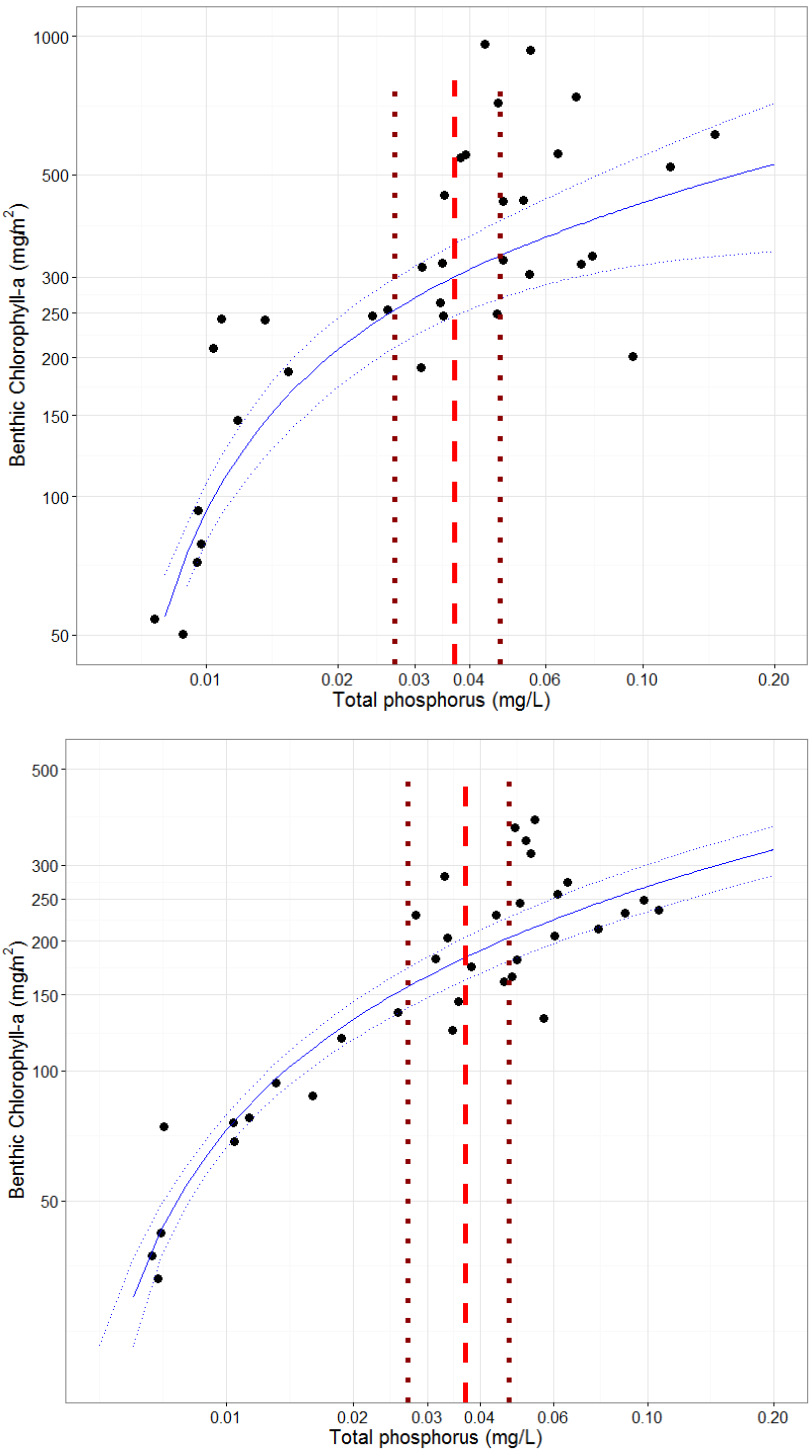

Figure 23. Mean benthic chlorophyll-a in year 1 (dry year, upper panel) and year 2 (wet year, lower panel) to annual mean total phosphorus. The fitted solid blue line is the result of a generalized additive model (GAM, deviance explained=88.5% and 89%, years 1 and 2 respectively) with 95% confidence limits shown as fine dotted lines around the fitted line. The mean chlorophyll-a values were weighted by inverse of the standard deviation to account for uncertainty. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

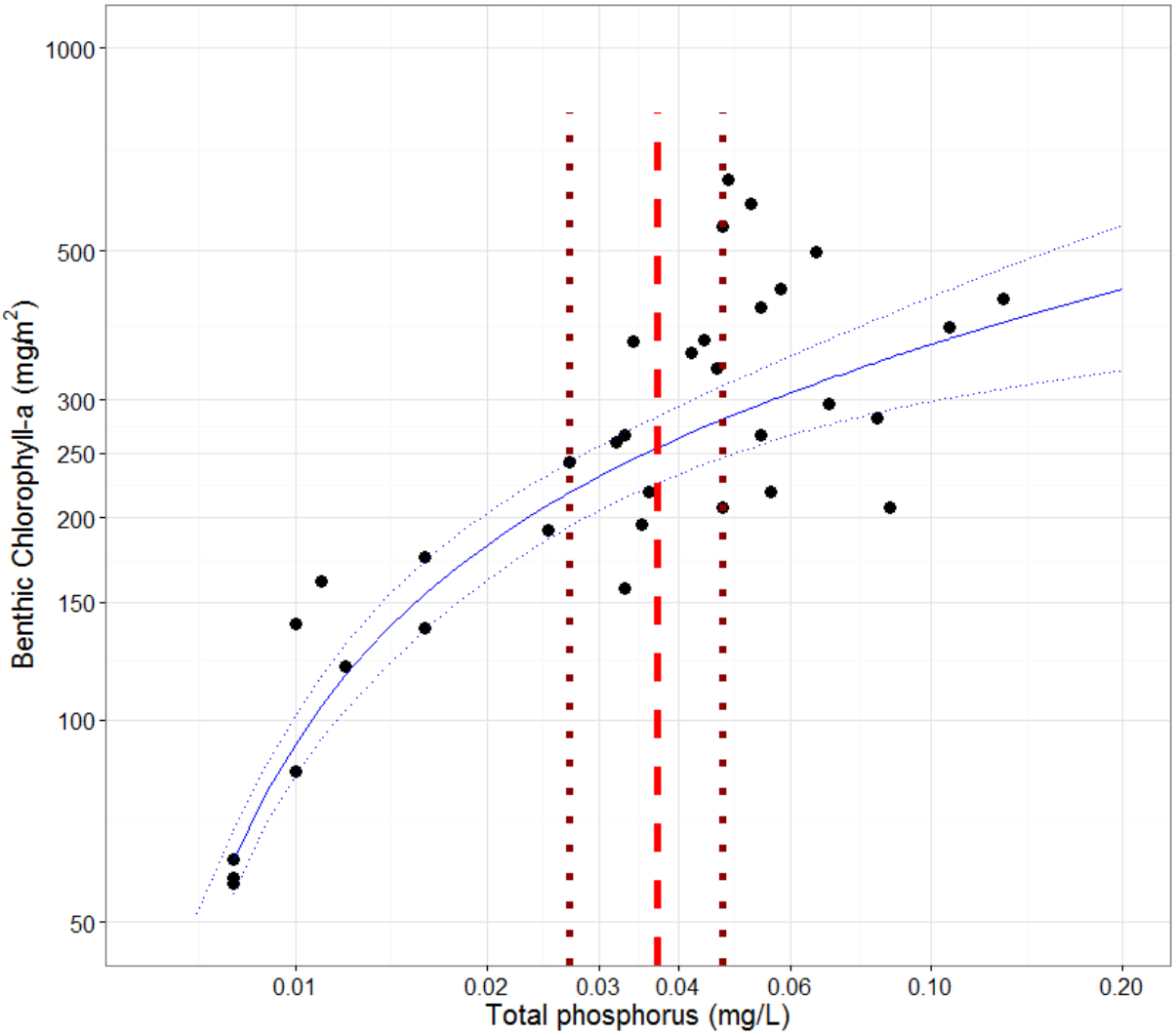

Figure 24. Mean benthic chlorophyll-a in response to mean 2-year total phosphorus. The fitted solid blue line is the result of a generalized additive model (GAM; deviance explained=90%, $p<0.00001$) with 95% confidence limits shown as fine dotted lines around the fitted line. The mean chlorophyll-a values were weighted by inverse of the standard deviation to account for uncertainty. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

Table 10. Predicted mean benthic chlorophyll-a in response year 1, year 2, and years 1 and 2-year mean total phosphorus at concentrations spanning 0.01 to 0.1 mg/L. The predictions are based on GAM models for each of the 3 data sets, with years 1 and 2 illustrated in the previous figure.

	Predicted benthic chlorophyll-a, mg/m^2								
	Year 1 (Dry, stable flows)			Year 2 (Wet, many storm flows)			Years 1 and 2, combined		
TP (mg/L)	Mean	5% CI	95% CI	Mean	5% CI	95% CI	Mean	5% CI	95% CI
0.010	93	80	107	73	67	79	92	83	102
0.020	209	173	244	131	118	145	182	161	203
0.027	255	211	298	157	140	174	218	194	242
0.030	270	224	317	166	148	184	230	205	256
0.037	300	247	354	183	163	204	254	225	283
0.040	311	255	368	190	168	212	263	232	293
0.047	333	269	397	204	180	227	280	246	314
0.050	342	275	409	209	184	233	287	251	323
0.060	366	289	444	224	197	251	307	265	348
0.075	396	304	488	243	213	273	331	281	381
0.100	435	319	550	267	234	301	362	298	427

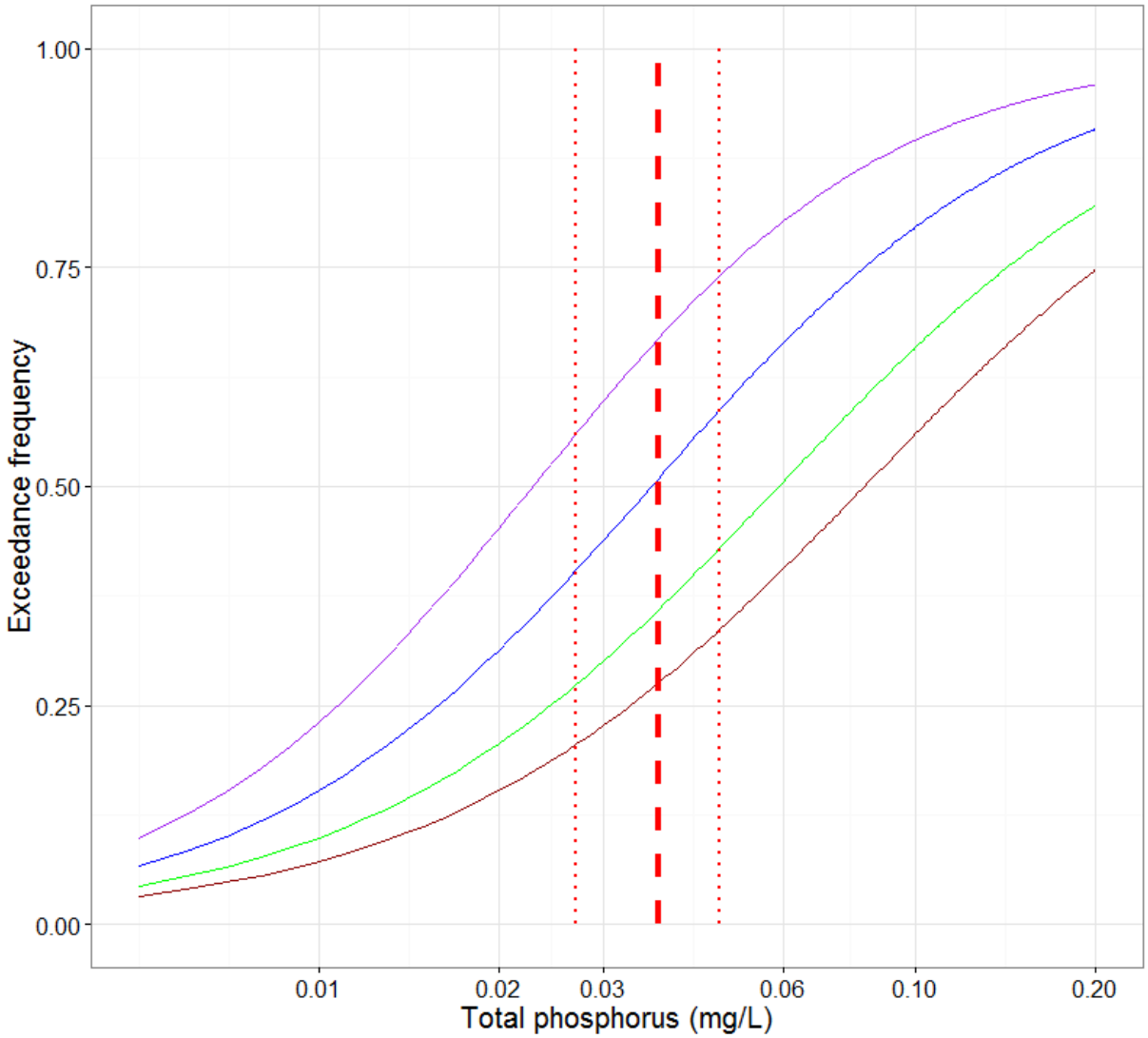

Figure 25. Exceedance frequencies of 150 (purple), 200 (blue), 250 (green), and 300 (dark red) mg/m^2 benthic chlorophyll-a in response to 2 year mean total phosphorus. GLM models for each response variable were fit using a binomial probability distribution and a logit link function. TP was a highly significant predictor in all 4 models ($p<0.00001$).

Table 11. Predicted exceedance frequencies of 150, 200, 250, and 300 mg/m^2 benthic chlorophyll-a for year 1 (dry year), year 2 (wet year), and years 1 and 2 combined in response to mean total phosphorus.

		Predicted exceedance frequencies of benthic chlorophyll-a vs. TP											
		150 mg/m2			200 mg/m2			250 mg/m2			300 mg/m2		
Year 1	TP (mg/L)	Mean	2.5%	97.%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
	0.010	0.29	0.17	0.41	0.25	0.14	0.35	0.10	0.04	0.17	0.12	0.05	0.19
	0.020	0.68	0.59	0.77	0.47	0.38	0.56	0.21	0.14	0.29	0.22	0.15	0.30
	0.027	0.81	0.74	0.88	0.58	0.50	0.65	0.28	0.21	0.35	0.29	0.22	0.36
	0.030	0.84	0.78	0.91	0.61	0.54	0.69	0.30	0.23	0.37	0.31	0.24	0.38
	0.037	0.90	0.84	0.95	0.68	0.61	0.75	0.36	0.29	0.43	0.36	0.29	0.44
	0.040	0.91	0.86	0.96	0.71	0.63	0.78	0.38	0.31	0.45	0.39	0.31	0.46
	0.047	0.94	0.90	0.98	0.75	0.68	0.82	0.43	0.35	0.51	0.43	0.35	0.51
	0.050	0.95	0.91	0.98	0.77	0.70	0.84	0.45	0.37	0.53	0.45	0.37	0.53
	0.060	0.96	0.94	0.99	0.81	0.74	0.88	0.50	0.41	0.59	0.50	0.41	0.59
	0.075	0.98	0.96	1.00	0.86	0.79	0.92	0.57	0.46	0.67	0.56	0.45	0.66
	0.100	0.99	0.98	1.00	0.90	0.84	0.96	0.65	0.53	0.77	0.64	0.51	0.76
Year 2													
	0.010	0.14	0.06	0.22	0.06	0.01	0.11	0.09	0.03	0.16	0.03	0.00	0.06
	0.020	0.29	0.20	0.37	0.16	0.09	0.24	0.20	0.12	0.28	0.08	0.03	0.13
	0.027	0.37	0.29	0.45	0.24	0.16	0.31	0.27	0.20	0.34	0.12	0.06	0.18
	0.030	0.40	0.33	0.48	0.27	0.19	0.34	0.30	0.22	0.37	0.14	0.08	0.20
	0.037	0.47	0.40	0.55	0.33	0.26	0.41	0.36	0.29	0.43	0.18	0.12	0.24
	0.040	0.50	0.42	0.57	0.36	0.29	0.44	0.38	0.31	0.45	0.20	0.14	0.26
	0.047	0.55	0.47	0.63	0.42	0.34	0.50	0.43	0.35	0.51	0.24	0.18	0.31
	0.050	0.57	0.49	0.65	0.45	0.36	0.53	0.45	0.37	0.53	0.26	0.19	0.33
	0.060	0.62	0.54	0.71	0.52	0.42	0.61	0.51	0.42	0.60	0.32	0.23	0.40
	0.075	0.69	0.59	0.78	0.60	0.49	0.71	0.58	0.48	0.69	0.40	0.28	0.51
	0.100	0.89	0.80	0.97	0.87	0.77	0.98	0.83	0.72	0.95	0.75	0.55	0.95
Years 1 and 2													
	0.010	0.23	0.09	0.37	0.15	0.04	0.27	0.10	0.00	0.20	0.07	0.00	0.15
	0.020	0.45	0.33	0.57	0.31	0.20	0.43	0.21	0.11	0.30	0.15	0.07	0.23
	0.027	0.56	0.44	0.68	0.40	0.31	0.50	0.27	0.18	0.37	0.21	0.13	0.28
	0.030	0.60	0.50	0.70	0.44	0.34	0.54	0.30	0.20	0.40	0.23	0.13	0.33
	0.037	0.67	0.57	0.77	0.51	0.41	0.61	0.36	0.26	0.46	0.28	0.18	0.37
	0.040	0.69	0.60	0.79	0.54	0.44	0.63	0.38	0.28	0.48	0.29	0.20	0.39
	0.047	0.74	0.64	0.84	0.59	0.49	0.69	0.43	0.31	0.55	0.34	0.24	0.43
	0.050	0.76	0.68	0.84	0.61	0.51	0.71	0.45	0.33	0.57	0.35	0.26	0.45
	0.060	0.80	0.72	0.88	0.66	0.55	0.78	0.51	0.39	0.62	0.41	0.29	0.52
	0.075	0.85	0.77	0.93	0.73	0.61	0.84	0.57	0.44	0.71	0.47	0.34	0.61
	0.100	0.90	0.82	0.97	0.80	0.68	0.91	0.66	0.50	0.82	0.56	0.38	0.74

Diel dissolved oxygen and pH

Multiprobe data sondes were deployed for 48-h at a minimum of 25 sites during August 2014 (near median baseflow conditions, late summer) and September 2015 (high baseflow conditions). The following figures illustrate the relationship between 6 month mean TP and minimum dissolved oxygen and maximum pH recorded during each 48-h deployment.

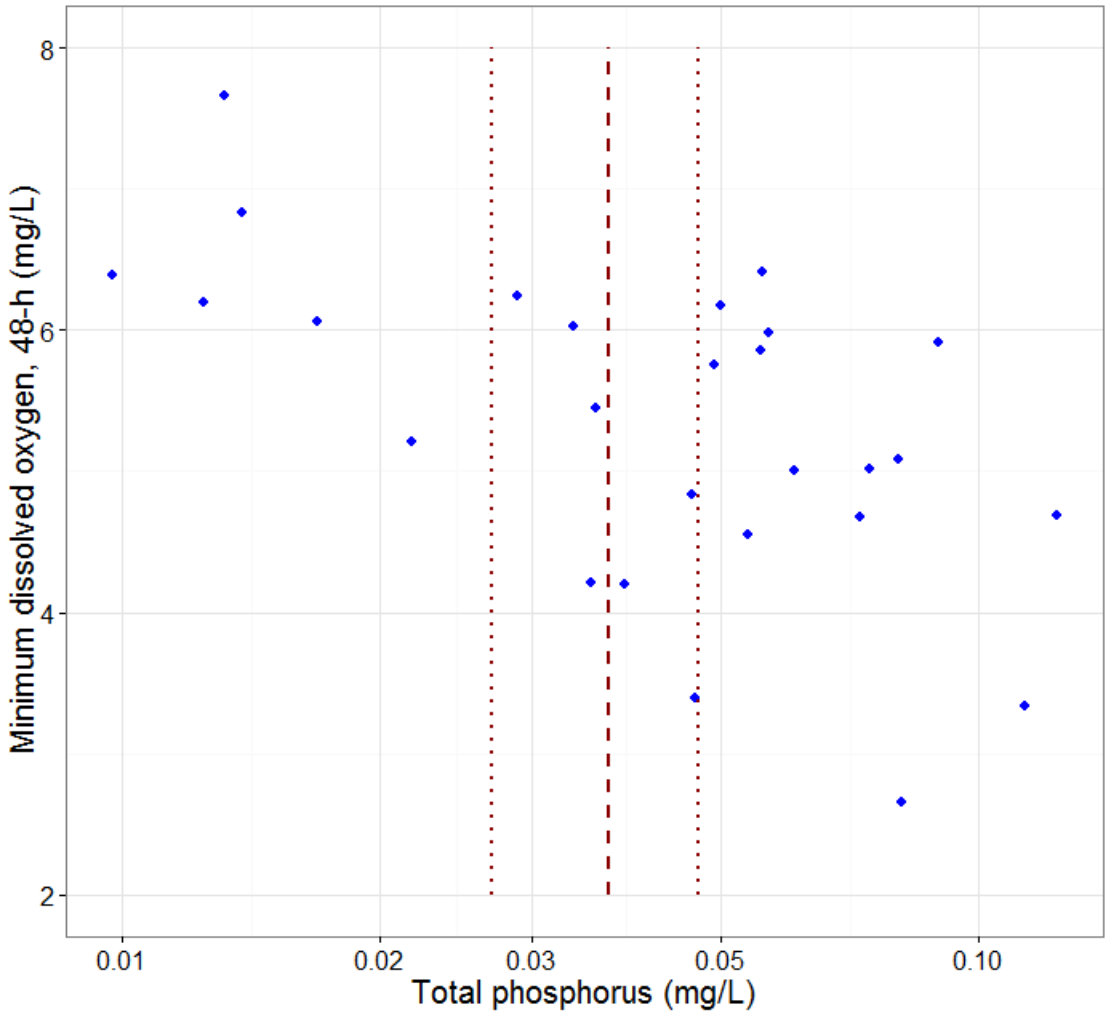

Figure 26. Minimum 48-h dissolved oxygen in August 2014 in response to mean 6-month total phosphorus. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

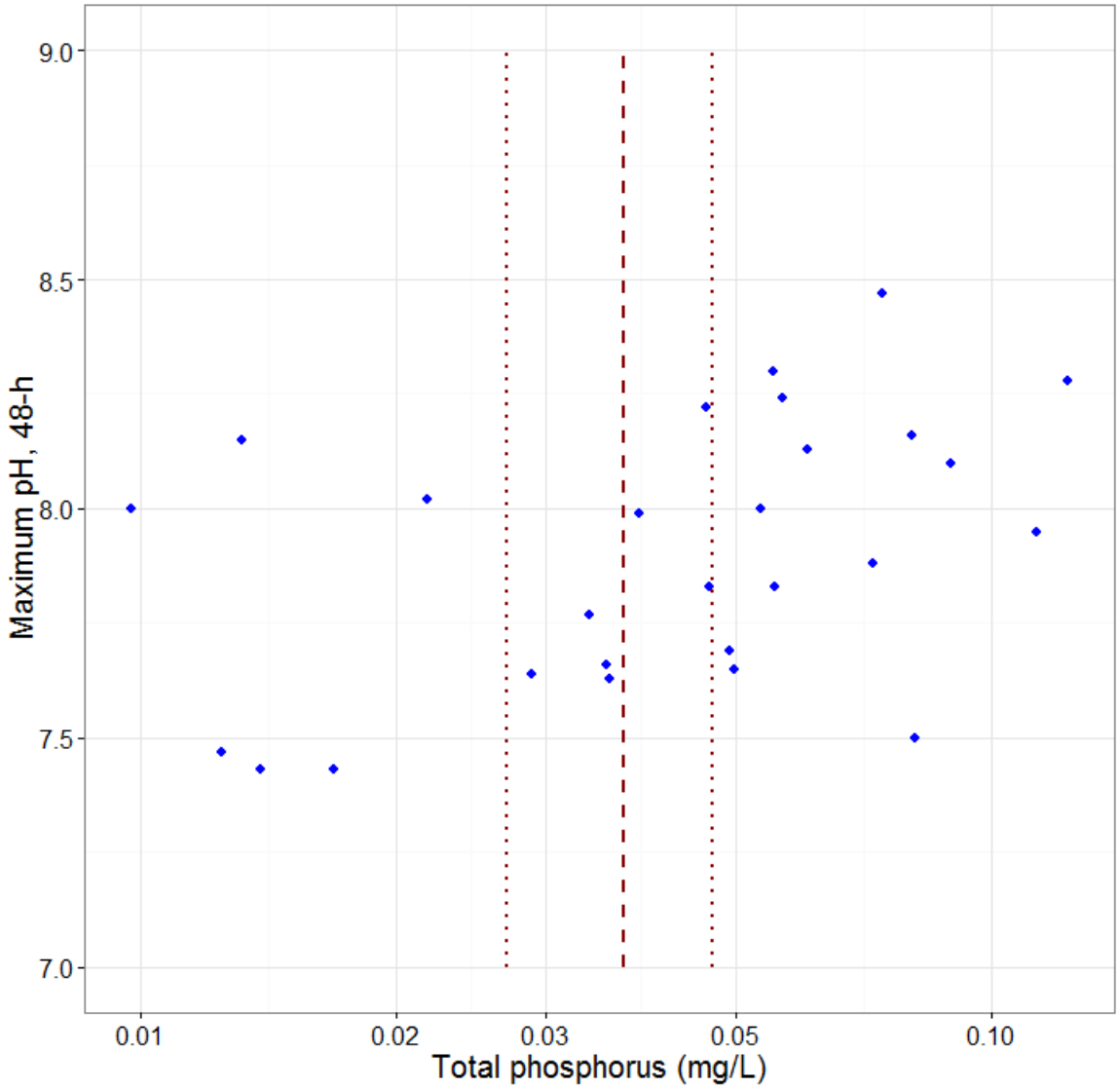

Figure 27. Maximum 48-h pH in August 2014 in response to mean 6-month total phosphorus. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

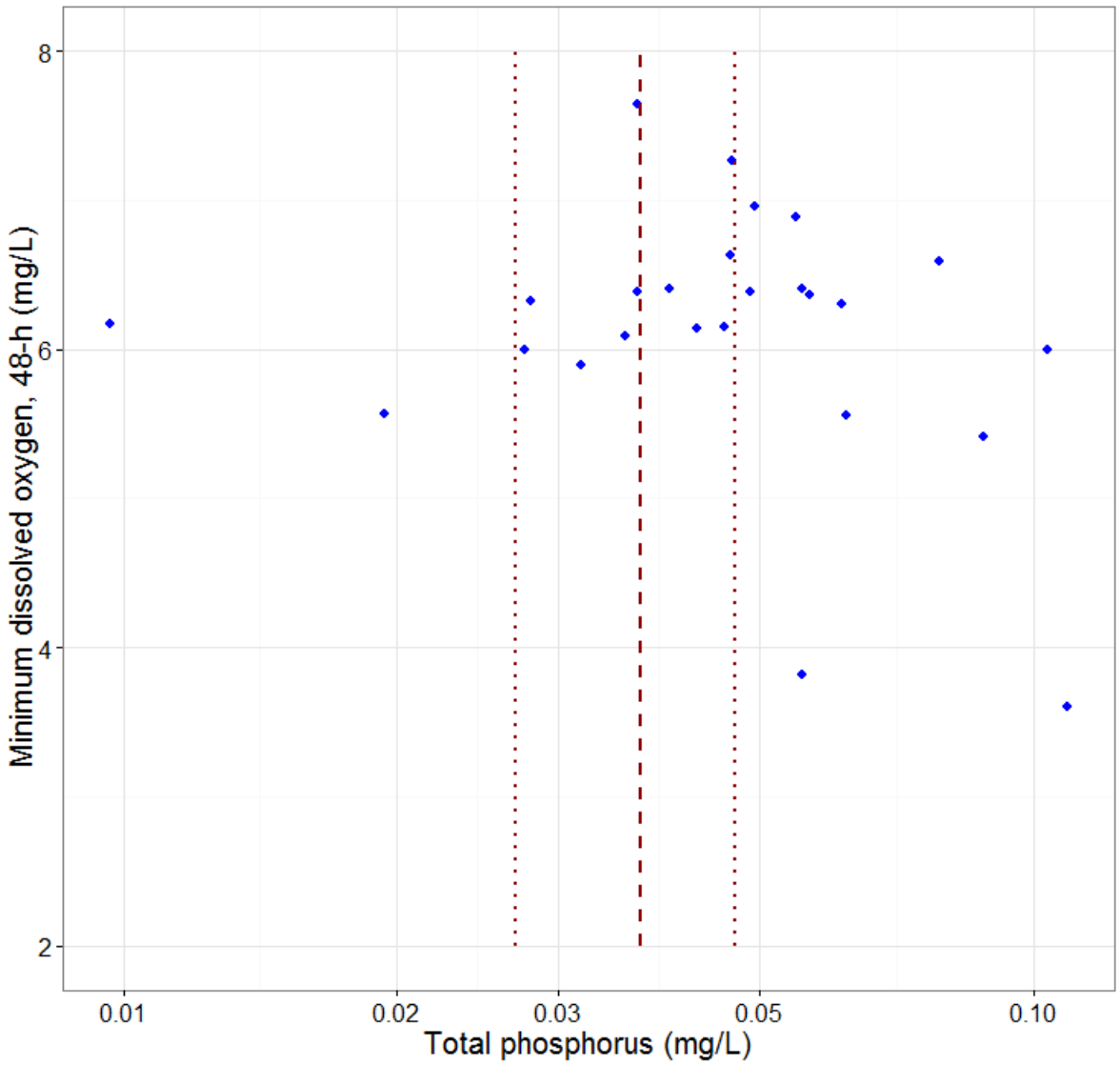

Figure 28. Minimum 48-h dissolved oxygen in September 2015 in response to mean 6-month total phosphorus. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

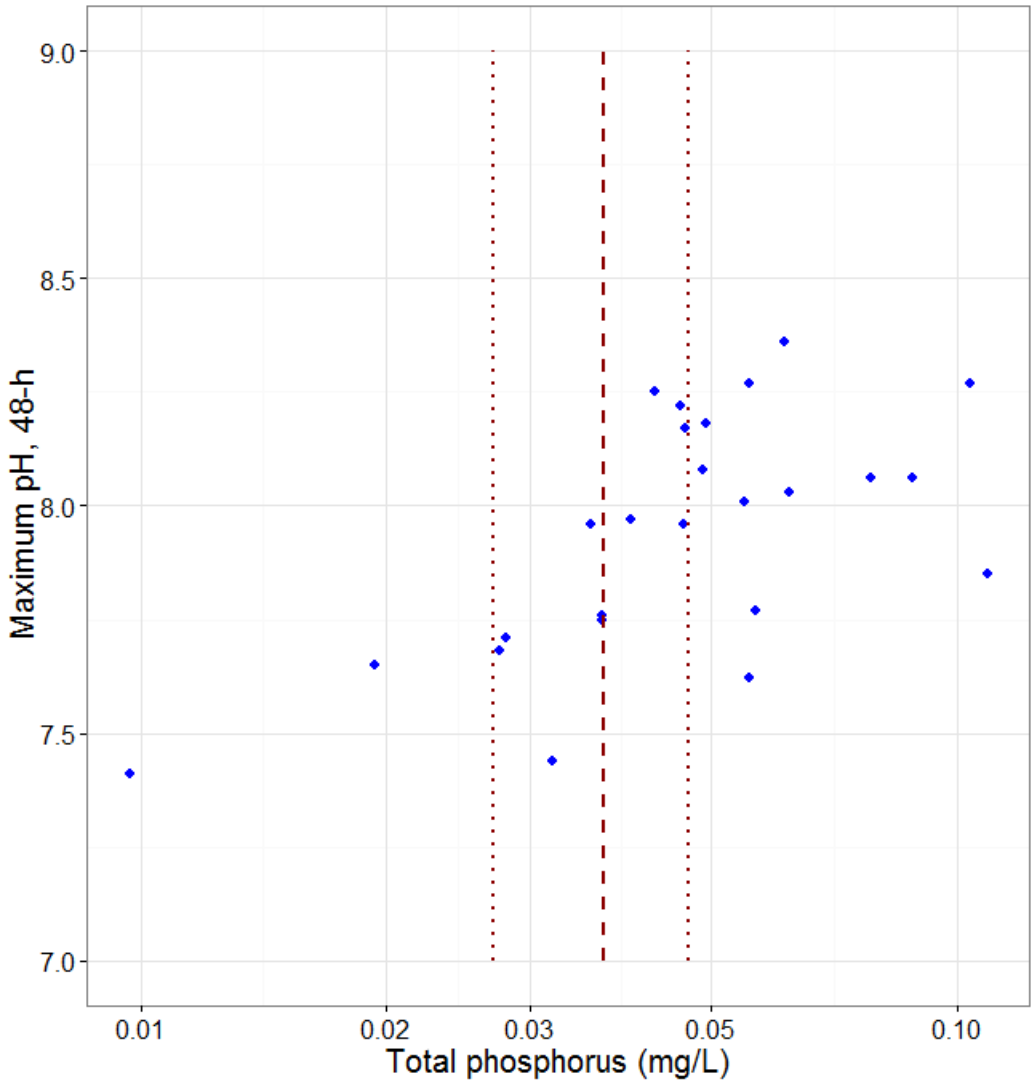

Figure 29. Maximum 48-h pH in September 2015 in response to mean 6-month total phosphorus. The red vertical dashed line corresponds to 0.037, whereas the dotted red vertical lines correspond to 0.027 and 0.047 mg/L, respectively.

Macroinvertebrates

The dominant grazing macroinvertebrate taxon in the study streams were snails in the family Pleuroceridae (Figure 30). Seasonally, pleurocerid densities varied considerably. Particular sites would have relatively few during one event but, by the next event, had exploded to levels such as those shown in the photograph.

Pleurocerids achieved densities up to 2000 individuals/m^2 based on estimates from Hess samples (Figure 31). The highest densities were observed in streams near the upper end of the phosphorus gradient, such as Spring Creek (AR; SPAR1), Osage Creek (OSAG1, OSAG2), Sager Creek (SAGE1), and Flint Creek (FLIN1, 2, and 3).

Figure 30. Photo of the stream bottom at Flint Creek (FLIN3) in late summer 2014, illustrating high densities of pleurocerid snails, the dominant algal grazing macroinvertebrate in the study streams.

Snails were abundant in at least a few streams during every event, regardless of season. Densities were likely underestimated during event 1 (June 2014) because snails would fall to the bottom of the stream bed when cobble and gravel were agitated to dislodge macroinvertebrates into the Hess sampler. Methods were adjusted during the following events such that rocks within the sampler were carefully lifted off the bottom and brushed directly into the net bag on the Hess sampler.

The only seasonal pattern evident was the nearly complete elimination of snails in events 11 and 12, which followed the historic flood of December 2015. This partially explains the very rapid growth of algae following the flood, as there was little to no grazing pressure by snails. Moreover, stonerollers (*Campostoma* spp.) were not actively grazing during the winter and early spring, thus February and April 2016 represented a nearly unrestricted growth response to nutrients.

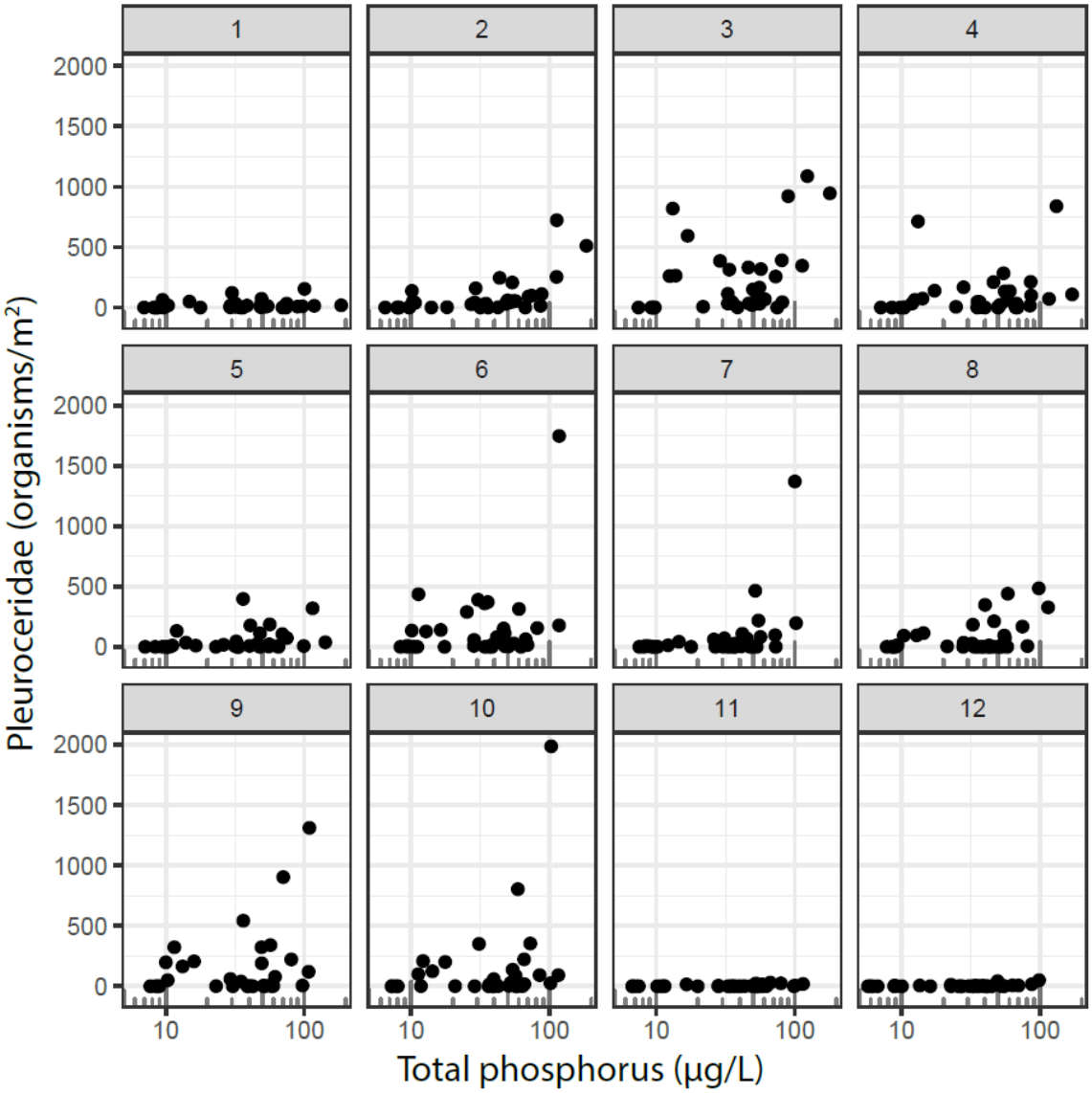

Figure 31. Densities of pleurocerid snails versus 6 month mean TP across the 2-year study period. Numbers in the upper panels are event numbers (1-12).

Responses of other macroinvertebrates varied but generally showed increases in density with increasing levels of TP. The following figure illustrates the mean response of each of the functional feeding groups of macroinvertebrates to TP over the 2-year study period.

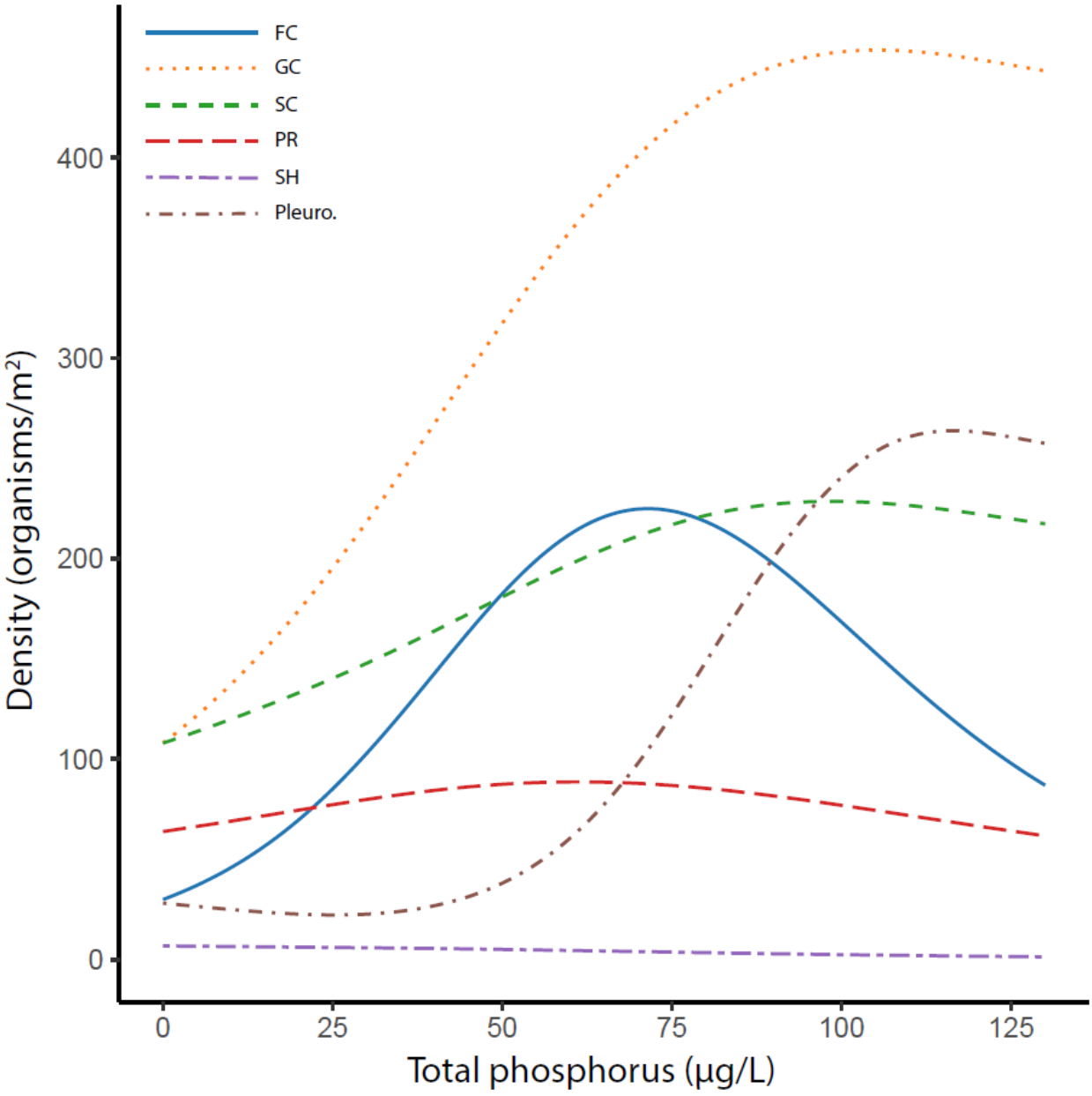

Figure 32. Mean responses of macroinvertebrate functional feeding groups to total phosphorus over the 2-year study period. FC=filtering collectors; GC=gathering collectors; SC=scrapers/grazers of algae, excluding pleurocerid snails; PR=predators; SH=shredders; Pleuro=pleurocerid snails.

Summary

The following histogram, which was requested by the Joint Study Committee, synthesizes the change points estimated by change point analysis and TITAN on all of the focal biological response variables analyzed using those techniques: benthic chlorophyll-a, *Cladophora* biovolume, nuisance taxa proportion, and community-level thresholds for negative and positive responding taxa (TITAN). Because analyses were conducted on several different TP durations, the 6 month duration was chosen for this summary because it was very similar to longer durations and was a stronger predictor than shorter durations in most cases.

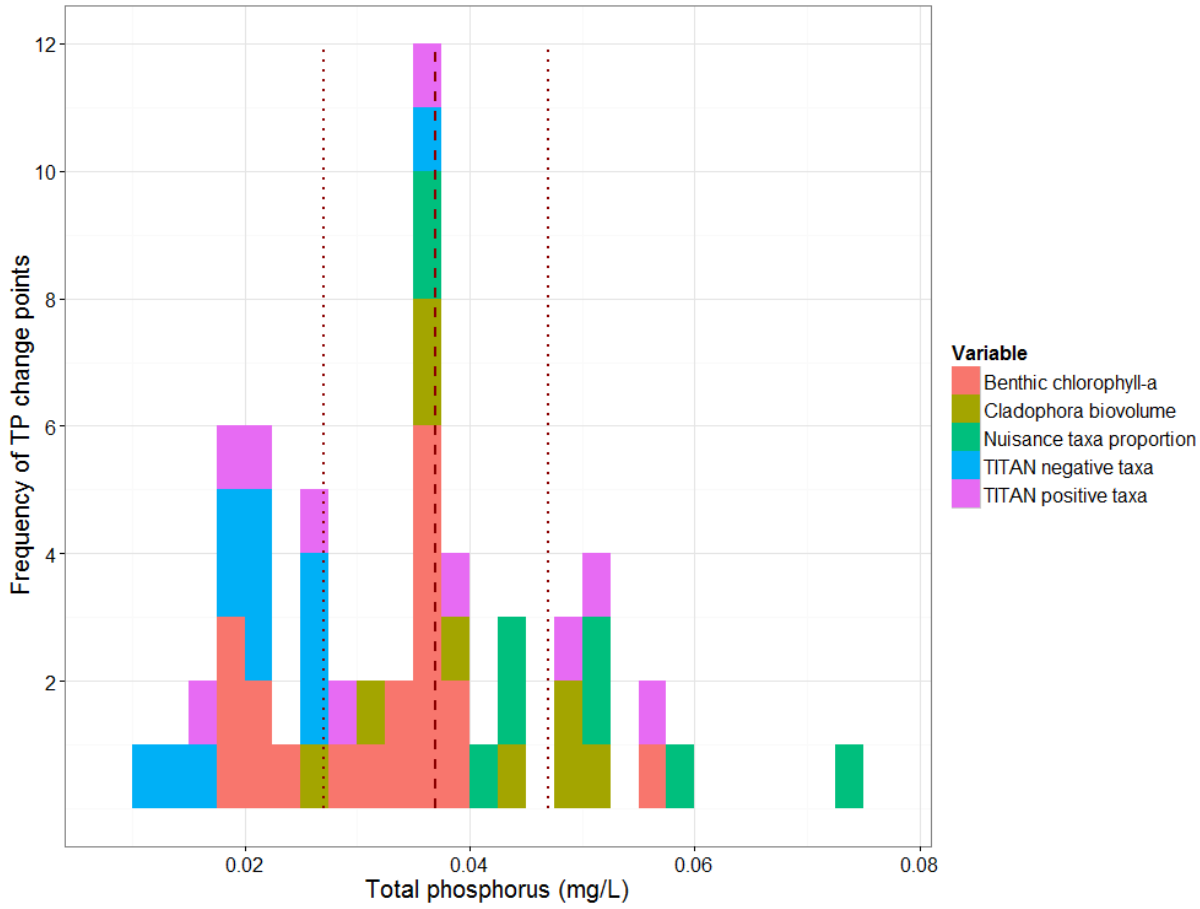

Figure 33. Histogram illustrating the distribution of total phosphorus change points across several response variables over the 2-year study period. Shown are change points associated with 6-month mean TP and instantaneous and mean responses that correspond to the TP data. The dashed red vertical line corresponds to 0.037, whereas the dotted vertical lines are 0.027 and 0.47, respectively.

Acknowledgments

This study was funded by the Arkansas-Oklahoma Joint Study Committee. Numerous individuals facilitated the execution of this study in accordance with Second Statement of Joint Principals and Actions and the approved work plan.

OWRB, OCC, and University of Arkansas provided historical data that helped guide site selection. Oklahoma Scenic Rivers Commission provided logistical support by showing the Baylor team access points on various rivers and helping us gain access on private lands. Days Inn of Tahlequah and Peyton's Place on the Illinois River provided greatly discounted accommodations to the Baylor team during field sampling trips that allowed the contractor to spend more time collecting data than would otherwise have been possible with the study budget.

Field sampling, sample processing, laboratory analyses, and data entry was completed by the outstanding efforts of Katherine Hooker, Morgan Bettcher, Stephen Elser, Stephen Cook, Caleb Robbins, and Lauren Housley. Dr. Jeffrey Back also participated in all aspects of the study, particularly performing water chemistry analyses and quality assurance oversight. Dr. Ryan King participated in all field sampling events, reviewed all data, performed data analyses and drafted the final report.

Drs. Stephen Porter and Barbara Winsborough performed taxonomic identifications and biovolume estimates of soft algae and diatoms, respectively.

Dr. Thad Scott served on the Joint Study Committee until early 2016. His service was greatly appreciated by the committee and the Baylor team.

Literature Cited

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

Baker, M.E., and R. S. King. 2013. Of TITAN and straw men: an appeal for greater understanding of community data. *Freshwater Science* 32:489-506

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers*. USEPA, Washington.

Biggs, B. J., & Kilroy, C. 2000. *Stream periphyton monitoring manual*. NIWA.

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

Breiman, L., 2001. Random forests. *Machine learning* 45:5-32.

De'ath, G. and Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.

Dodds, W.K. and Gudder, D.A., 1992. The ecology of *Cladophora*. *Journal of Phycology* 28:415-427.

Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H., Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D. and Poff, N.L., 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1-13.

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. & Richardson C.J. (2003) Integrating bioassessment and ecological risk assessment: An approach to developing numerical water-quality criteria. *Environmental Management* 31:795-809.

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 watershed urbanization. *Ecological Applications*. 21:1659-1678.

King, R.S. and M.E. Baker. 2014. Use, misuse, and limitations of Threshold Indicator Taxa Analysis (TITAN) for natural resource management. pp 231-254 In: G. Guntenspergen (editor), *Application of Threshold Concepts in Natural Resource Decision Making*, Springer.

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

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

Taylor, J. M., J. A. Back, and R. S. King. 2012. Grazing minnows increase benthic autotrophy and enhance the response of periphyton elemental composition to experimental phosphorus additions. *Freshwater Science* 31:451-462.

Taylor, J. M., R.S. King, A. Pease, and K.O. Winemiller. 2014. Nonlinear response in stream ecosystem structure to low level phosphorus enrichment. *Freshwater Biology* 59:969-984.

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

USEPA (United States Environmental Protection Agency). 1998. *National strategy for the development of regional nutrient criteria*. EPA 822-R-98-002. Office of Water, Washington, DC

USEPA. 2010. *Using Stressor-response Relationships to Derive Numeric Nutrient Criteria*.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society* 24:706–723.

Zuur, A. F. 2009. *Mixed effects models and extensions in ecology with R*. Springer.

Zuur, A.F., Ieno, E.N. and Elphick, C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1:3-14.