10-1-2016

Database Analysis to Support Nutrient Criteria Development (Phase II)

B. E. Haggard

M.A. Evans-White

L.B. Massey

E.M. Grantz

ARKANSAS WATER RESOURCES CENTER | PUBLICATION MSC382
FUNDED BY THE TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

DATABASE ANALYSIS TO SUPPORT NUTRIENT CRITERIA DEVELOPMENT (PHASE II)

2016 October

ARKANSAS WATER RESOURCES CENTER

UofA DIVISION OF AGRICULTURE RESEARCH & EXTENSION University of Arkansas System

Arkansas Water Resources Center | Publication MSC382
Funded By the Texas Commission on Environmental Quality

The intent of this publication of the Arkansas Water Resources Center is to provide a location whereby a final report on water research to a funding agency can be archived.

The Texas Commission on Environmental Quality (TCEQ) contracted with University of Arkansas researchers for a multiple year project titled "Database Analysis to Support Nutrient Criteria Development".

This publication covers the second of three phases of that project and has maintained the original format of the report as submitted to TCEQ. This report can be cited either as an AWRC publication (see below) or directly as the final report to TCEQ.

Haggard, B.E., J.T. Scott, M.A. Evans-White, L.B. Massey, and E.M. Grantz. 2016. Database Analysis to Support Nutrient Criteria Development (Phase II). Arkansas Water Resources Center, Fayetteville, AR, MSC Publication 382: 368 pp.

DATABASE ANALYSIS TO SUPPORT NUTRIENT CRITERIA DEVELOPMENT

Final report submitted in fulfillment of
Contract number 582-12-21325 to:

Julie Mcentire
Project Manager, Water Quality Standards Group
Texas Commission on Environmental Quality

By

Principal Investigators:
Brian E. Haggard, Professor and Director, Arkansas Water Resources Center
J. Thad Scott, Assistant Professor, Crop, Soil and Environmental Sciences Department
Michelle Evans-White, Assistant Professor, Biological Sciences Department
University of Arkansas

Project Management Team:
Leslie B. Massey, Program Associate, Arkansas Water Resources Center
Erin M. Grantz, Program Associate, Crop, Soil and Environmental Sciences Department

August 30, 2013

Table of Contents

List of Tables

List of Figures

Section 1: Streams and Rivers

E.M. Grantz, L.B. Massey, B.E. Haggard, J.T. Scott

EXECUTIVE SUMMARY

The Clean Water Action Plan, released in 1998 by the United States Environmental Protection Agency (USEPA), established a national set of nutrient criteria for the 14 aggregate ecoregions across the United States, directing states and tribes to adopt these criteria or pursue development of scientifically defensible criteria at the state level. For streams and rivers, the two main approaches for criteria development focus on the frequency distribution of median concentrations of a general population or a select group of sites representing reference conditions and statistical analysis of stressor-response relationships between nutrients and biological response variables. Predictive approaches have focused on establishing relationships between nutrient concentrations and algae, macroinvertebrates, and fish communities.

The objective of Section 1 was to provide statistical support to the Texas Commission on Environmental Quality (TCEQ) to assist in the development of numeric nutrient criteria for Texas streams and rivers by TCEQ. The first step in this process was to compile geospatial, water quality, and bioassessment data from 2,273 stations spanning 23 basins across Texas. These data were provided by TCEQ and collected under non-biased conditions. Following data reorganization and reduction, median values for each parameter were estimated at each station with 10 observations or greater and compiled into a median database. The parameters of primary concern were total phosphorus (TP), ortho-phosphate (PO_4-P; SRP), total nitrogen (TN), nitrate plus nitrite N (NO_x-N), and sestonic chlorophyll-a (chl-a). Frequency distributions, including the minimum, 10th, 25th, 50th, 75th, and 90th percentiles, and maximum values of these parameters, were calculated for the general population and selected reference sites at multiple spatial scales, specifically, by basin, eco-region levels III and IV, and basin by level III eco-region. Frequency distributions are presented in Section 1.1 and were intended to provide TCEQ with the percentile estimates for Texas rivers and streams recommended by the USEPA for setting nutrient and chl-a criteria.

States are progressing to development of nutrient criteria, but questions remain regarding the legitimacy of promulgating one numeric criterion across areas that may contain multiple basins, various eco-regions, and a myriad of land uses. Section 1.2 provides analyses of potential geospatial variability in total nutrient (TP and TN), chl-a, and Secchi transparency for Texas streams and rivers using classification and regression tree (CART) and non-parametric changepoint analysis (nCPA). Geospatial variables included land use/land cover (LULC) categories, permitted municipal wastewater treatment (WWTP) plant discharge, and regions (basin, ecoregion III, and basin by ecoregion III). Changepoints were identified in several geospatial predictor variables to describe variability in nutrient parameters, particularly median TP concentrations. Models for biological response variables were generally weak. Permitted municipal WWTP discharge weighted by watershed area was directly related to total nutrients, and a threshold of 0.031 mgd/km^2

explained >30% of variability in median TP concentrations. This WWTP discharge threshold was was used in subsequent stressor-response analysis to group “low” and “high” WWTP discharge stations. Among the categorical region variables, total nutrients, particularly median TP concentrations, could most effectively be grouped by basin by ecoregion III areas.

The frequency distribution approach should be used in conjunction with other statistically based methods that evaluate stressor-response relationships in aquatic systems. Section 1.3 provides analyses of potential nutrient thresholds (TP, TN, NO_x-N, NH_4-N, and SRP) to biological response (chl-a measured spectrophotometrically and fluorometrically (chl-a spec and chl-a fluoro), Secchi transparency, and 24-hour dissolved oxygen (DO) flux) in Texas streams and rivers. Median parameter estimates were divided by station into 3 datasets (“all” stations, “low” WWTP discharge, and “high” WWTP discharge) based on geospatial analysis resulst of total nutrients vs. area-weighted permitted municipal WWTP discharge. The total nutrient thresholds identified for biological response variables ranged from 0.063 to 0.11 mg/L TP and 0.70-1.1 mg/L TN. Nutrient thresholds identified for Secchi transparency were consistly lower than those identified for chl-a, particularly for chl-a spec. Total nutrient thresholds identified for the “all” and “low” datasets were consistently identical or similar, indicating that stations potentially highly impacted by permitted municipal WWTP discharge did not drive the analysis of the “all” stations dataset. No statistically significant relationships between DO Flux and total nutrients were found for any of the datasets, but DO Flux was directly related to chl-a concentrations. CART models based on total nutrients, especially TP, declined in strength and thresholds increased in magnitude in the high flow dataset in comparison with the all stations and low flow datasets.

A subset of Texas streams and rivers has undergone more intensive biological and habitat sampling, in addition to water quality data collection. Section 1.4 provides CART and nCPA analyses of stressor-response relationships in these streams, adding indices of biotic integrity as response variables and habitat as a stressor variable to previously considered parameters. Median indices of biological integrity, including the Fish Index of Biotic Integrity (Fish IBI) and the macrobenthic rapid bioassessment index of biotic integrity (RBIBI), as well as habitat scores (HQI) were integrated with median nutrient concentrations across several temporal periods, including the total period of record, annual, index, and critical periods. Individual IBI’s and habitat scores were also paired with point water quality measurements collected simultaneously or within the same week to month as the biological and habitat data. For the fish and macrobenthic indices of biotic integrity (IBIs) considered in this study, habitat quality (HQI) was consistently a strong predictor variable with threshold values ranging from 18-21. Nutrient thresholds for these variables were consistently weak or not statistically significant. Model strength and significance did increase for Fish IBI vs. both TN and TP and RBIBI vs. TP as the temporal scale of analysis widened from the paired observations to the period of record dataset. This finding may reflect the fact that bioassessment metrics such as Fish IBI and RBIBI were implemented as indicators of long-term nutrient concentrations and habitat conditions in streams that are potentially superior to point measurements of nutrient concentrations.

INTRODUCTION

The Clean Water Action Plan, released in 1998 by the United States Environmental Protection Agency (USEPA), established a national set of nutrient criteria for the 14 aggregate ecoregions across the United States, five of which are located partly within Texas. These numerical values were set for both causative (e.g., nutrients) and response (e.g., chlorophyll and transparency) variables which are associated with the prevention and assessment of eutrophic conditions in streams and rivers. However, local and regional influences can affect water quality, resulting in median nutrient concentrations and biological conditions that differ from the USEPA recommendations (e.g., Ice et al., 2003; Smith et al. 2003; Binkley 2004, Longing and Haggard 2010; Evans-White et al., 2013). Therefore, states, tribes and others may choose to adopt the criteria set by the USEPA or establish scientifically defensible nutrient criteria for streams and rivers specific to local areas of concern. Two commonly accepted statistical approaches to developing criteria include percentile analysis of data frequency distributions and stressor-response relationships.

The frequency distribution method does not require prior knowledge of individual stream conditions to set criteria; rather, the criteria are developed relative to the population of streams and rivers in a specific area (e.g., state, basin or ecoregion). The USEPA (2000) has suggested two statistical methods to identify nutrient criteria based on percentile analysis of data frequency distributions. The first method establishes the 75th percentile of a data distribution of reference or minimally impacted streams and rivers a criterion; the second is based upon the 25th percentile of the general population. The USEPA (2000) suggests that both approaches should result in similar criterion (Figure 1.1). However, studies have shown that the estimated criterion can be highly variable between these approaches (Suplee et al. 2007 and Herlihy and Sifeneos 2008), and generally 75th percentile estimates have been less conservative than 25th percentile estimates (Evans-White et al. 2013). This could indicate that the aggregate ecoregions are too coarse a scale for establishing nutrient criteria. Therefore, the basin or smaller ecoregion level might be more appropriate (Rohm et al. 2002). In addition, the 75th percentile approach is constrained by the limited existence of true reference condition streams (Dodds and Oaks, 2004). Nonetheless, the frequency distribution method is a tool that can aid states, tribes and other groups when setting nutrient criteria.

Stressor-response approaches, as recommended by USEPA (2010), evaluate biological conditions over a range of environmental gradients including nutrients and habitat quality. Regression tree models have been used to explain the variance in stream nutrient concentrations as a function of land use, ecoregion, and other watershed attributes (e.g., Herlihy and Sifneos 2008). Classification and regression tree (CART) analysis is an empirical modeling technique and is useful for identifying ecological thresholds and hierarchical structure in predictor variables (De'ath and Fabricius 2000). CART uses recursive partitioning to divide data into subsets that are increasingly homogeneous, invoking a tree-like classification that can explain relationships that may be difficult to reconcile with conventional linear models (Urban 2002). Categorical variables (e.g., station location, basin, ecoregion or land-use classifications) may also be used as independent variables in CART analysis, which provides another advantage to using CART rather than traditional regression techniques. CART and other similar methods have been used to identify thresholds

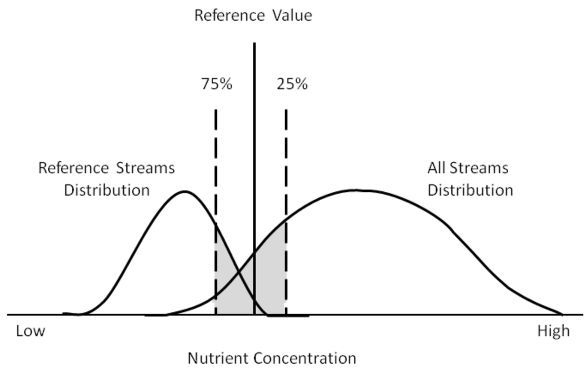


Figure 1.1. Distribution of data collected from reference condition streams and the general stream population and the associated percentile distribution used to develop nutrient criteria.

and hierarchical structure in environmental correlates of various biological processes in aquatic ecosystems (King et al. 2005, East and Sharfstein 2006). King et al. (2005) used CART specifically to identify thresholds in nutrient concentrations which resulted in shifts in ecological structure and function in the Florida Everglades region.

The objectives of this section are:

1) to discuss the frequency distribution of median nutrient concentrations and response variables for Texas streams and rivers calculated from data acquired from the Texas Commission of Environmental Quality (TCEQ) at various spatial scales including basins, level III ecoregions, level IV ecoregions and basin-level III ecoregion combinations;
2) to explore the relationship between median nutrient concentrations (focusing on TP and TN), as well as common biological parameters (focusing on Secchi transparency and chlorophyll-a) and watershed attributes (both numeric and categorical) for Texas streams and rivers, providing a defensible approach from which Texas streams and rivers could be grouped by watershed attributes;
3) to identify nutrient threshold values associated with changes in the magnitude or variability of commonly measured biological response variables for Texas streams and rivers;
4) to identify nutrient and habitat threshold values associated with changes in the magnitude or variability of commonly measured bioassessment variables for Texas streams and rivers.

The methods used to identify thresholds in biological response variables relative to nutrient stressor variables in Texas streams and rivers are supported by available literature on the development of nutrient criteria for rivers and streams (e.g., King et al. 2005; East and Sharfstein 2006). Other states that are in the process of developing area-specific nutrient criteria have considered the results of multiple statistical approaches, as discussed, before selecting criteria levels. Several states have adopted site-specific criteria for streams and rivers, but only Wisconsin, Florida, New Jersey, and Vermont have statewide criteria. These criteria are presented in Table 1.1 as summarized by Evans-White et al., 2013.

Table 1.1. Summary of numerical nutrient (i.e., total nitrogen; TN and total phosphorus; TP) criteria for streams and rivers in water quality standards (WQS) and the year they were publish across the 48 conterminous states (taken from Evans-White et al., 2013).

State	Local	TN mg/L	Local	TP mg/L	WQS Year
Arizona	site-specific	0.50-1.00	site-specific	0.05-0.20	2010
California	site-specific	--	site-specific	--	--
Florida	statewide	0.67-1.87	statewide	0.06-0.49	2012
Georgia	--	--	site-specific	--	2012
Montana	site-specific	0.13-1.36	site-specific	0.01-0.12	2012
Nevada	site-specific	1.5-2.9 (2.4-4.0)	site-specific	0.10-0.33 (0.05-0.10)	2012
New Jersey	site-specific	2.0	statewide	0.10	2011
New Mexico	--	--	site-specific	0.10	2012
New York	site-specific	--	--	--	2002
Oklahoma	--	--	site-specific	0.04	2012
Oregon	--	--	site-specific	0.07	2012
Vermont	statewide	0.2-5.0	site-specific	0.01	2012
Washington	--	--	site-specific	0.03	2012
Wisconsin	--	--	statewide	0.07-0.10	2012

1.1. STREAMS AND RIVERS DATABASE DEVELOPMENT, MEDIAN CALCULATION, AND FREQUENCY DISTRIBUTIONS

Methods

Water Quality Database

Data Acquisition, Compilation and Reduction. TCEQ provided a database of water quality data collected from 1968 to 2012 from freshwater streams and rivers throughout Texas. The collected data was from 2,273 stations spanning 23 watersheds and was divided among three Microsoft Excel workbooks. The data described 116 stream characteristics and water quality parameters including nutrients, sediments, transparency, physico-chemical parameters, as well as others.

For the purposes of advanced statistical analyses conducted during this project, only data collected under specific monitoring code types (as decided by TCEQ) and from 2000 to 2010 was used. Therefore, the database was sorted and any data collected before calendar year 2000 or after 2010 was removed. Data collected under the monitoring code type Biased Flow (BF) was also removed since data collected under this circumstance were not necessarily representative of baseline water quality conditions. The data received from TCEQ were output to a single column format within the files, so the data were reorganized into a useable format. The data were sorted by Basin ID and a new Microsoft Excel worksheet was created for each individual basin. Each basin worksheet was then restructured using the pivot table function in Microsoft Excel so that each parameter and the associated data were unique to an individual column; a portion of this process was accomplished with a Mircosoft Excel Macro (see Appendix 1.1 for Excel Macro code). Any estimated data points (i.e., those reported with a < or >) were flagged and used in the database without the associated qualifying sign. The data were flagged using a Microsoft Excel Macro (Appendix 1.2).

Several additional parameters were calculated from the original data provided. Nitrate plus nitrite-nitrogen (NO_x-N) and total nitrogen (TN) were calculated if the necessary N species were provided by TCEQ in the original data file. In addition, diel change (i.e., 24 hour maximum minus 24 hour minimum) was calculated for dissolved oxygen, temperature, conductivity, pH, and turbidity. The additional parameters were added to each station worksheet using a Microsoft Excel Macro (Appendix 1.3).

Due to the volume of data provided, several parameters were removed from the median database because of lack of data and duplication of parameters, or because TCEQ indicated that the parameter could be removed from the database.

Median and Frequency Distribution Calculations. For this study, frequency distribution and, subsequently, stressor-response analyses were conducted on station medians in order to focus on broadly applicable regional and statewide trends. Because each stream and river in Texas was not equally represented in the raw water quality dataset, conducting statistical analyses on medians removes potential site-specific bias for sites that are over- or under-represented in the raw dataset. Furthermore, biological response and nutrient stressor data did not always overlap in the raw data. Conducting analyses with median values allowed comparison of long-term trends in biological and nutrient data for these stations. Medians were calculated for each Station ID using a Microsoft Excel Macro (Appendix 1.4). Median values were calculated based on at least 10 data points, i.e. no medians were calculated if less than 10 data points were available for a given parameter at a given station. The calculated medians for each Station ID were then compiled into one database using a Microsoft Excel Macro (Appendix 1.5). This database was merged with the GIS and LULC data and used in advanced statistical analysis.

TCEQ also provided a list of 75 stations that represented least disturbed stream conditions, and these sites spanned 14 basins and 10 level III ecoregions across the state. The calculated medians associated with these sites were compiled into a separate database from which frequency distributions of the medians were calculated.

Frequency distributions (minimum value, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles and maximum value) for water quality parameters TP (TCEQ parameter code 00665), TN (calculated parameter code 00600C; TCEQ parameter code 00625 + 00630, 00625 + 00593 or 00625 + 00615 + 00620), NO_x-N (calculated parameter 00630C; TCEQ parameter code 00630, 00593 or 00615 + 00620), PO_4-P (TCEQ calculated parameter code 00671C; TCEQ parameter code 00671 or 70507), and sestonic chlorophyll-a (chl-a) measured fluorometrically (TCEQ parameter code 70953; chl-a fluoro) were calculated using Microsoft Excel. For this study, a parameter combining chl-a measured spectrophotometrically (parameter code 32211; chl-a spec) and chl-a fluoro was not created due to inconsistencies between the methods (Laurie Eng, personal communication). Data were more complete and censorship was less of a concern for chl-a fluoro than for chl-a spec. Spectrophotometric chl-a data were commonly censored at a relatively high detection limit

(10 µg/L). Analysis exploring the effects of censored data on chl-a spec median calculation in Texas reservoirs indicated that, when censored data exceeded 16% of the raw data for a station, chl-a medians (i.e. the 50th percentile) were increasingly overestimated when the detection limit was substituted for censored observations as the level of censoring increased because the median could not be calculated to be a value below the detection limit (see Section 2.3). This censored data effect would have been apparent when considering low percentiles in the medians frequency distribution, and the 25th percentile estimate would have frequently been equal to the detection level. Therefore, frequency distributions for sestonic chl-a were only calculated for the fluorometric method in this study. Frequency distributions were calculated for the general streams population at multiple spatial scales including basin, level III ecoregion, basin by level III ecoregion (i.e., unique combinations of basin and level III ecoregions combined), and level IV ecoregion. Frequency distributions were also calculated for the least disturbed stations.

Geospatial Database

A geospatial database contained within a Microsoft Excel file was provided by TCEQ that identified land use and land cover data for the water quality stations located on streams included in this study. The geospatial descriptors were provided for the drainage basin and riparian area. Drainage boundaries for the monitoring stations were limited to the upstream extent of the Subbasin (HUC-8) boundaries in the USGS Watershed Boundary Dataset, which may not include the total watershed boundary from the monitoring station to the headwaters (Archuleta et al. 2012). The descriptors included percent open water, developed-open, developed-low intensity, developed- medium intensity, developed-high intensity, barren land, deciduous forest, evergreen forest, mixed forest, shrub/scrubland, grassland/herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands. These descriptors were reduced to five categories including percent developed (i.e., open, low intensity, medium intensity, barren land), forest (i.e., deciduous, evergreen, mixed and shrubland), agriculture (i.e., grassland/herbaceous, pasture/hay, cultivated crops), developed plus agriculture (i.e., open, low intensity, medium intensity, barren land, grassland/herbaceous, pasture/hay, cultivated crops) and wetlands (i.e., woody and emergent herbaceous wetlands). Additional geospatial information for each site was provided including drainage area, slope, municipal discharges, basin ID, level III ecoregion ID, and level IV ecoregion ID.

Data Quality Assurance and Control

Data quality checks were employed frequently throughout the database reorganization and data calculation processes. The original source files were maintained in an unaltered form, and subsequent changes to each database were saved under unique file names. Data transferred from one file to the next were checked for accuracy by comparing first and last rows and the row count between files. In addition, when calculations were preformed, including manual calculations and those calculated using Microsoft Excel Macros, at least 10 percent of calculations were checked for accuracy following the secondary data quality assurance project plan (QAPP).

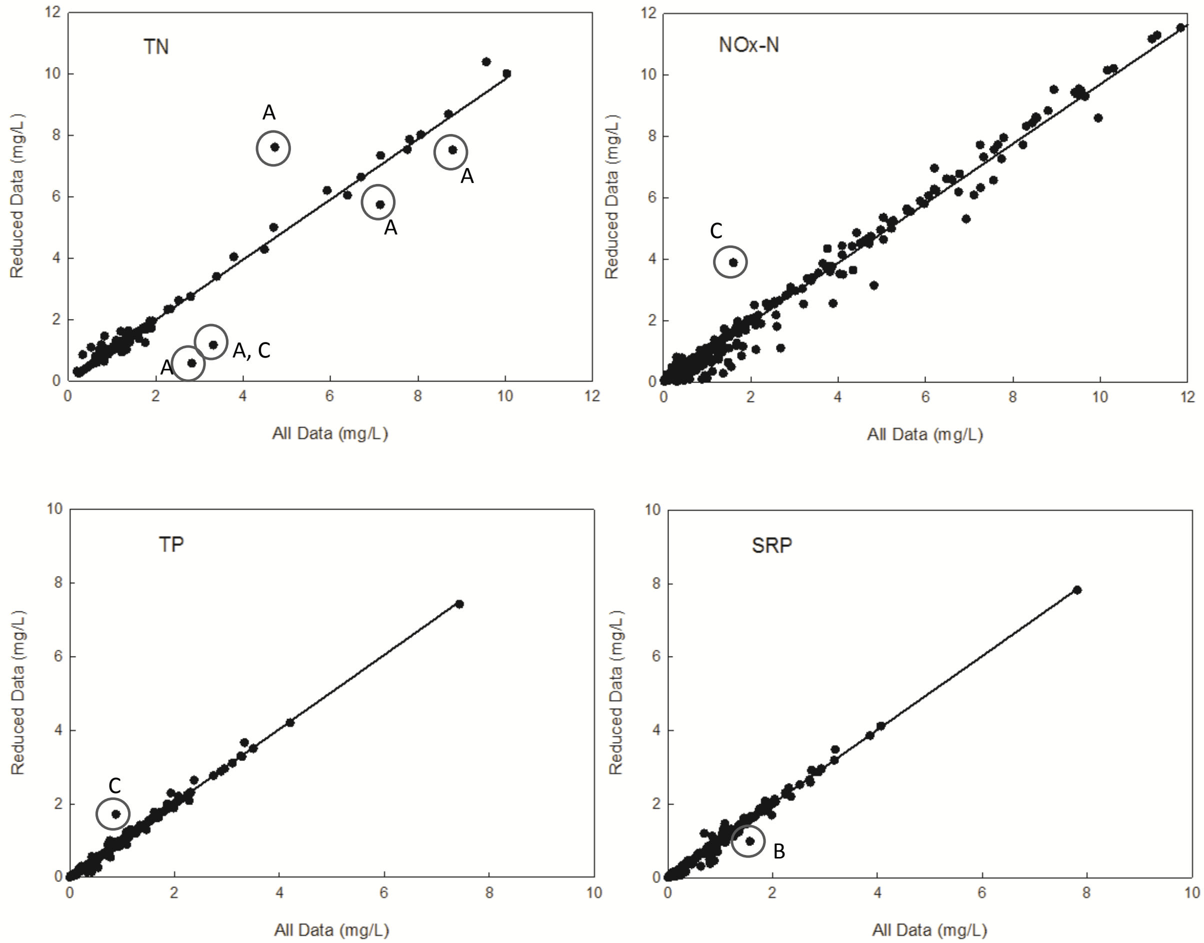
Results and Discussion

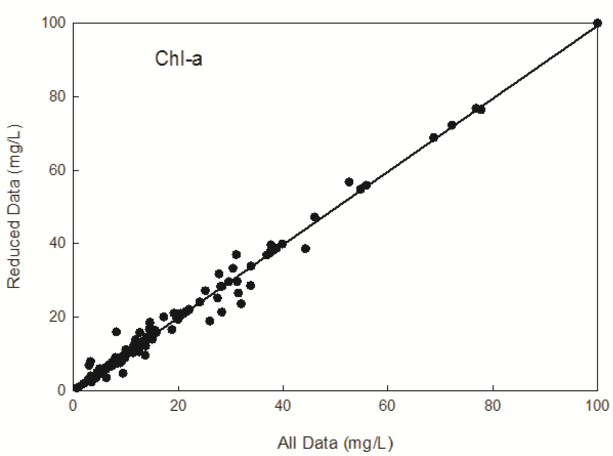
Additional Data and Reductions

The data for FY 2012-13 were specifically selected by TCEQ based on the project objectives by which the data were collected (i.e., monitoring type code), and data collected under the monitoring codes Biased Season (BS), Ecoregion Study (ER), Routine (RT), Special Study (SS), SWQM Acquired Routine/Baseline Water Sampling (XR), and Diel Sampling (DI) were used in the FY 2012 data analysis. The data used previously in FY 2011 analysis were compiled from all monitoring type codes, including in addition to those previously listed, Biased Event (BE), Biased Flow (BF), Citizen Monitoring (CM), Intensive/Systemic-sub-watershed Monitoring (IS), Special Event Monitoring (SE), and TMDL QAPP-305(b)/305(d) Assessment (TQ). Data collected under these monitoring type codes were collected under biased or non-representative conditions and potentially affected estimation of central tendencies for monitoring stations. Therefore these data were removed from the water quality database for FY 2012-13 analyses. In order to evaluate whether biased or non-representative data influenced median calculation during the FY 2011 project, we compared FY 2011 medians calculated using all the data and FY 2012-13 medians calculated using the reduced data (as outlined above) for the previously listed parameters of interest. In most cases, we observed similar median values for each station ID. Relating FY 2011 and FY 2012-13 medians using simple linear regression showed that most values were similar between project years, and that the regression slopes were consistently near one (Slope=0.97-1.00, $R^2 \geq 0.96$, $P<0.0001$; Figure 1.1.1). However, for a few stations, medians differed noticeably between project years.

In several cases, fewer data were available for a station in the FY 2012-13 database compared to the FY 2011 database, because a large proportion of the data for that station were collected under a biased or non-representative monitoring code type and were not included in FY 2012-13. Removing data collected under biased or non-representative monitoring type codes (specifically BF, IS, and TQ) from the database therefore affected median calculation for these stations (Figure 1.1.1). For example, the FY 2011 median TP value for Station 17406 considered 21 TP concentrations that were collected from 2001-2003 under the IS monitoring code and one TP concentration collected in 2009 during a biased flow event. As a result of removing these data, the median TP concetnration at this site increased from 0.88 mg/L (n=107) in FY 2011 to 1.71 mg/L (n=85) in FY 2012-13.

Other observed differences in medians were due to the addition of data (e.g., additional parameter codes or sampling dates) in the FY 2012-13 database compared to the FY 2011 database. For example, SRP median concentrations for Station 12911 decreased from 1.57 mg/L (n=13) in FY 2011 to 0.98 mg/L (n=16) in FY 2012, because data from three additional samples collected from July to November 2010 were provided in the FY 2012 dataset (Figure 1.1.1). TCEQ also provided additional NO_x-N data under the parameter code 00593 in FY 2012-13; in FY 2011 this code was considered to be only NO_3-N, and was not used in NO_x-N calculations because of the abundance of other NO_3-N data collected under the parameter codes 00618 and 00620.





Parameter	Equation	R^2	P Value
TN	y=0.98x+0.025	0.96	<0.0001
NO_x-N	y=0.97x-0.007	0.99	<0.0001
TP	y=1.00x-0.002	0.99	<0.0001
SRP	y=1.00x-0.003	0.99	<0.0001
Chl-a	y=0.99x+0.034	0.99	<0.0001

Figure 1.1.1. Relationship between median nutrient and chlorophyll-a concentrations calculated from all available data (FY 2011, previous project) and data selected based on collection methods (i.e., reduced data, FY 201213) for Texas streams and rivers. (A) Influenced by parameter code priority when calculating new parameters; (B) Influenced by additional data provided in FY 2012-13; (C) Influenced by less data provided in FY 2012 (due to select monitoring codes).

The priority or approach by which new parameters were calculated also had an effect on the parameter medians in FY 2011 compared to FY 2012-13. This was particularly evident in the comparison of TN data between project years (Figure 1.1.1). Since measured TN concentrations were limited in the data provided by TCEQ, we calculated TN based on the data availability of species of nitrogen (e.g., dissolved and organic) which could be summed to calculate TN. The priority of parameter codes by which TN was calculated varied from FY 2011 to FY 2012-13; in FY 2011 we calculated NO_X-N values, TKN values, and then TN values. The process was simplified in FY 2012 because of the additional data provided y the NO_X-N code 00593 under, which resulted in fewer calculated NO_X-N values. In addition, we did not calculate additional values for TKN in FY 2012-13, and we simply used the TKN data provided as parameter code 00625.

Finally, we observed differences in sestonic chl-a data between FY 2011 and FY 2012-13 because in FY 2011 we created a sestonic chl-a parameter which combined data measured spetrophotometrically and fluorometrically; whereas in FY 2012, we only considered sestonic chl-a measured fluorometrically. However, in most cases the differences between FY 2011 and 2012-13 median chl-a ranges was minimal (Table 1.1.1). The only instance in which the maximum or minimum chlorophyll-a value changed by more than 20% was for Level IV Ecoregions. The range of TP, TN, NO_X-N, SRP, and fluorometric chl-a station medians observed at the Basin, Level III Ecoregion, Level IV Ecoregion, and BasinXLevel III Ecoregion scales in both FY 2011 and 2012-13 are provided in Table 1.1.1

Table 1.1.1. Range of 25th percentile median nutrient and chlorophyll-a concentrations calculated from all available data (FY 2011, previous project) and data selected based on collection methods (i.e., reduced data, FY 2012-13) for Texas streams and rivers.

Parameter	Aggregate Ecoregion	Project Year	Basin	Level III Ecoregion	Level IV Ecoregion	BasinXLevel III Ecoregion
Total P (mg/L)	0.010-0.13	FY 2011	0.050-0.30	0.030-0.17	0.020-0.37	0.015-0.80
	--	FY 2012-13	0.050-0.55	0.050-0.17	0.022-0.37	0.019-0.73
Total N (mg/L)	0.12-0.88	FY 2011	0.40-4.7	0.31-1.2	0.26-3.0	0.30-4.7
	--	FY 2012-13	0.44-2.1	0.35-1.6	0.27-2.5	0.31-6.5
Nitrate-N (mg/L)	--	FY 2011	0.040-3.1	0.030-0.28	0.020-1.6	0.020-3.1
	--	FY 2012-13	0.040-2.7	0.040-0.69	0.020-0.98	0.020-6.0
Soluble Reactive P (mg/L)	--	FY 2011	0.040-0.12	0.020-0.040	0.004-0.21	0.018-1.0
	--	FY 2012-13	0.013-0.29	0.020-0.080	0.008-0.15	0.013-3.1
Chlorophyll-a (µg/L)	0.93-3.0	FY 2011[1]	1.5-31	3.00-11.4	0.06-27.5	1.47-30.5
	--	FY 2012-13[2]	1.9-33	3.00-9.81	3.00-20.8	1.86-33.3

[1]Chlorophyll-a parameter in FY 2011 combined spectrophotometric and fluorometric methods.
[2]In FY 2012-13 frequency distributions were calculated for fluorometric chl-a only.

General Stream Conditions Frequency Distributions

Frequency distributions of nutrient and chl-a median concentrations were calculated at multiple spatial scales, and the distributions discussed here represent general stream conditions. Frequency distributions calculated for basin and level III ecoregion are presented following, and distributions calculated for basin-level III ecoregion and level IV ecoregion are presented in Appendix 1.6.

Basin. The State of Texas is divided into 23 basins (Appendix 1.7) which are categorized as river (65%) or coastal (35%) basin waters. River basin waters are the surface inland waters comprising the major streams and their tributaries while coastal basin waters are surface inland waters that discharge or in some way interconnect with bays or the Gulf of Mexico. The 25th percentile of median TP concentrations was less than 0.10 mg/L for 60% of the basins in Texas, and the 25th percentiles at these basins ranged from 0.05 to 0.08 mg/L (Table 1.1.2). The basins with less data tended to have 25th percentile median concentrations that were greater than 0.10 mg/L. For example, six of the eight basins where the 25th percentile was greater than 0.10 mg/L had 15 or fewer medians for distribution analysis, while the USEPA recommends a minimum of 30 data points be used when analyzing frequency distributions to guide nutrient criteria development (USEPA, 2000). The 25th percentile of median concentrations of PO_4-P data followed a pattern similar to that observed for TP, and the 25th percentiles of the medians of these parameters were strongly correlated ($R^2 = 0.94$; $p < 0.0001$). Basin 16, Lavaca River Basin, and Basin 22, Nueces-Rio Grande Costal Basin, had a 25th percentile median PO_4-P concentrations greater than 0.10 mg/L, but only five and eight median data points contributed to the frequency distribution at these basins, respectively. The 25th percentile PO_4-P concentrations ranged from 0.01-0.08 mg/L for the other basins (Table 1.1.2).

Fewer TN medians were available for analysis compared to other measured parameters, so frequency distributions were only calculated for 65% of the basins (Table 1.1.1). Furthermore, only five basins (i.e., Trinity River Basin (8), Brazos River Basin (12) and Colorado River Basin (14) San Antonio River Basin (19) and Rio Grande Basin (23) had more than 30 median data points contributing to the frequency distribution of the data; the 25th percentile of median TN concentrations ranged from 0.44 to 1.61 mg/L at these five basins. The range in the 25th percentile of median TN concentrations across all basins was from 0.40 to 2.05 mg/L. The 25th percentile of median NO_x-N concentrations ranged from 0.04 to 2.70 mg/L including the 10 basins which had fewer than 30 data points contributing to the median. The 25th percentile of the median concentrations of TN and NO_X-N were positively correlated ($R^2 = 0.36$; $p = 0.010$).

The 25th percentile chl-a data distribution was calculated for 70% of the basins and ranged from 1.8-5.0 µg/L, except for the Red River Basin (2) and the Nueces-Rio Grande Coastal Basin (22) which exhibited 25th percentile concentrations that were greater than those observed for the other basins (Table 1.1.2). Similar to the patterns observed for the frequency distribution of the other parameters, these basins had nine or fewer medians from which the distribution was calculated. The 25th percentile of the median chl-a concentrations were positively correlated to nutrient concentrations (TP: $R^2 = 0.61$; $p = 0.0003$; PO_4-P: $R^2 = 0.66$; $p = 0.0001$; TN: $R^2 = 0.18$; $p = 0.0969$; NO_X-N: $R^2 = 0.61$; $p = 0.0004$).

Table 1.1.2. Frequency distribution of median nutrient and chlorophyll-a concentrations for streams and rivers among basins in Texas, 2000-2010; distributions are based on the reduced data with select monitoring type codes excluded. *No data were available for Basin 17.

Total Phosphorus (TP; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	16	0.060	0.063	0.078	0.088	0.130	0.250	0.508
2	63	0.020	0.050	0.060	0.120	0.218	0.912	4.200
3	15	0.060	0.074	0.110	0.210	0.469	0.900	1.680
4	31	0.023	0.050	0.080	0.100	0.140	0.190	1.290
5	30	0.060	0.060	0.060	0.100	0.130	0.170	0.246
6	72	0.060	0.060	0.060	0.098	0.170	0.259	3.300
7	6	0.110	--	0.162	0.175	0.218	--	0.345
8	108	0.029	0.060	0.060	0.080	0.191	0.978	2.880
9	2	0.125	--	--	0.138	--	--	0.150
10	152	0.040	0.101	0.159	0.908	1.290	1.797	3.285
11	37	0.080	0.096	0.150	0.270	0.620	0.770	0.980
12	152	0.040	0.060	0.060	0.102	0.283	0.980	7.430
13	11	0.070	0.131	0.179	0.196	0.223	0.240	0.390
14	121	0.020	0.025	0.060	0.060	0.080	0.300	2.230
15	1	--	--	--	0.370	--	--	--
16	10	0.085	0.099	0.174	0.205	0.228	0.301	0.310
18	66	0.007	0.022	0.050	0.050	0.064	0.220	1.710
19	71	0.020	0.060	0.060	0.188	0.694	0.891	2.205
20	2	0.020	--	--	0.650	--	--	1.240
21	33	0.002	0.050	0.060	0.060	0.135	0.157	0.323
22	8	0.092	0.154	0.545	0.718	0.815	1.111	1.415
23	63	0.005	0.051	0.060	0.090	0.226	0.392	0.790
24	7	0.100	0.100	0.115	0.140	0.403	0.645	0.660

Total Nitrogen (TN; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	7	0.40	0.45	0.51	0.71	1.56	2.79	4.19
2	21	0.72	0.81	0.98	1.37	1.66	5.11	6.18
3	15	0.85	0.90	1.00	1.21	2.50	8.28	10.95
4	27	0.59	0.64	0.72	0.90	1.06	1.21	6.46
5	22	0.78	0.82	0.90	1.15	1.30	1.60	1.64
6	13	0.56	0.64	0.80	0.83	0.93	0.97	1.24
7	3	1.24	--	--	1.63	--	--	2.23
8	69	0.47	0.74	0.88	1.16	1.89	8.16	13.70
9	1	--	--	--	1.32	--	--	--
10	15	0.62	0.82	2.05	5.27	5.80	6.16	7.02
11	5	0.59	--	0.94	1.54	1.71	--	3.44
12	93	0.37	0.80	1.14	1.74	3.31	7.22	82.50
13	8	1.12	1.22	1.28	1.40	1.51	1.55	1.62
14	83	0.24	0.31	0.44	0.84	1.44	2.34	9.77
15	1	--	--	--	1.44	--	--	--
16	0	--	--	--	--	--	--	--
18	16	0.49	0.51	0.63	1.14	1.65	1.86	7.47
19	68	0.34	1.01	1.61	2.78	7.01	9.01	10.40
20	0	--	--	--	--	--	--	--
21	17	0.46	0.51	0.84	1.14	5.93	7.44	16.30
22	5	1.49	--	1.95	5.09	6.64	--	8.06
23	46	0.56	0.87	0.99	1.29	1.85	2.85	3.68
24	0	--	--	--	--	--	--	--

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	15	0.04	0.04	0.06	0.25	2.01	4.63	6.86
2	45	0.02	0.04	0.08	0.21	0.57	3.24	9.34
3	15	0.04	0.08	0.13	0.20	1.26	7.19	10.22
4	28	0.04	0.04	0.05	0.09	0.28	0.42	5.81
5	32	0.06	0.08	0.10	0.15	0.18	0.29	0.74
6	67	0.04	0.05	0.06	0.18	0.44	0.89	8.94
7	6	0.05	--	0.08	0.15	0.28	--	0.76
8	73	0.04	0.12	0.23	0.41	1.29	6.55	13.05
9	2	0.04	--	--	0.17	--	--	0.30
10	50	0.04	0.04	0.04	0.22	1.77	4.42	12.39
11	41	0.02	0.04	0.09	0.27	1.08	2.13	3.92
12	189	0.04	0.08	0.11	0.32	1.06	3.21	70.60
13	10	0.04	0.04	0.22	0.24	0.37	0.46	0.47
14	158	0.02	0.02	0.06	0.21	0.68	1.99	12.56
15	1	--	--	--	0.50	--	--	--
16	7	0.08	0.13	0.19	0.23	0.28	0.50	0.81
18	54	0.03	0.16	0.26	0.54	0.83	1.51	11.70
19	70	0.04	0.37	1.12	2.27	6.19	7.99	9.53
20	2	0.04	--	--	1.21	--	--	2.34
21	33	0.02	0.02	0.10	0.30	1.10	6.11	15.95
22	8	0.15	0.20	2.70	3.92	4.39	5.37	5.64
23	60	0.04	0.04	0.15	0.34	0.69	1.29	1.73
24	7	0.04	0.04	0.04	0.04	0.04	0.04	0.05

Ortho-Phosphate (PO_4-P; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	16	0.020	0.040	0.040	0.043	0.071	0.135	0.360
2	59	0.020	0.020	0.040	0.050	0.084	0.402	3.860
3	11	0.040	0.055	0.060	0.060	0.100	0.660	0.725
4	26	0.010	0.010	0.013	0.055	0.060	0.060	0.920
5	33	0.040	0.040	0.040	0.040	0.050	0.117	0.865
6	64	0.010	0.025	0.040	0.050	0.070	0.147	2.775
7	5	0.030	--	0.060	0.060	0.070	--	0.206
8	119	0.008	0.020	0.040	0.040	0.068	0.812	2.210
9	2	0.050	--	--	0.065	--	--	0.080
10	152	0.010	0.041	0.080	0.773	1.259	1.907	3.480
11	40	0.040	0.040	0.069	0.225	0.380	0.661	0.780
12	241	0.003	0.020	0.040	0.040	0.120	0.485	8.010
13	11	0.040	0.070	0.082	0.101	0.120	0.130	0.250
14	145	0.020	0.020	0.020	0.040	0.060	0.233	1.860
15	1	--	--	--	0.180	--	--	--
16	5	0.065	--	0.120	0.120	0.140	--	0.145
18	17	0.020	0.026	0.040	0.040	0.040	0.060	0.745
19	59	0.020	0.020	0.040	0.134	0.712	0.959	2.240
20	2	0.020	--	--	0.595	--	--	1.170
21	27	0.030	0.040	0.040	0.050	0.060	0.094	0.205
22	8	0.040	0.061	0.291	0.392	0.498	0.718	0.970
23	58	0.006	0.040	0.040	0.060	0.074	0.190	0.570
24	8	0.030	0.037	0.040	0.068	0.125	0.299	0.390

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	5	3.00	--	3.28	4.78	7.06	--	18.9
2	9	3.00	4.82	19.4	37.9	38.6	39.1	39.9
3	13	3.00	3.00	3.00	4.10	7.10	21.0	28.4
4	13	3.00	3.00	3.00	3.06	4.66	4.98	5.12
5	0	--	--	--	--	--	--	--
6	13	3.00	3.00	3.00	8.39	12.0	21.9	41.3
7	2	16.5	--	--	28.1	--	--	39.6
8	15	3.00	3.03	3.76	10.2	12.8	18.4	21.1
9	1	--	--	--	3.00	--	--	--
10	14	3.00	3.00	3.62	5.51	8.18	9.36	10.1
11	5	3.00	--	3.00	3.00	3.39	--	15.8
12	113	3.00	3.00	3.30	5.57	13.9	21.7	72.2
13	7	0.730	0.936	1.86	3.00	3.42	4.36	5.71
14	57	2.00	3.00	5.00	5.00	8.10	23.0	76.8
15	1	--	--	--	9.31	--	--	--
16	0	--	--	--	--	--	--	--
18	10	3.00	3.00	3.00	3.00	3.09	3.73	4.50
19	16	3.00	3.00	3.00	3.00	3.18	4.88	13.7
20	2	3.00	--	--	5.00	--	--	5.00
21	24	2.00	3.00	3.00	5.00	9.04	17.4	31.8
22	5	20.1	--	33.3	37.9	76.4	--	100
23	27	3.00	3.00	3.00	7.70	19.8	26.0	56.7
24	0	--	--	--	--	--	--	--

Level III Ecoregion. Texas is divided into 11 level III ecoregions comprised of deserts (9%), tablelands (9%), timbers (9%), plateaus (9%), prairies (9%), and plains (55%) (Appendix 1.8). The 25th percentile of median TP concentrations was calculated for all level III ecoregions (Table 1.1.3), except for High Plains where only three medians were available. The 25th percentiles spanned a similar range to that observed by basin, where most (90%) of the 25th percentiles of median TP concentrations were 0.10 mg/L or less (i.e, 0.05-0.10 mg/L). However, the Western Gulf Coastal Plain had a 25th percentile of 0.17 mg/L TP. The 25th percentiles of the median PO_4-P concentrations were less varied and ranged from 0.02 to 0.08 mg/L.

The ecoregions East Central Texas Plains and Western Gulf Coastal Plains had the highest 25th percentiles of median TN concentrations, which were 1.62 and 1.34 mg/L, respectively (Table 1.1.3). The 25th percentile of median TN concentrations were less than 1.0 mg/L for the remaining level III ecoregions, with the lowest value observed at Edwards Plateau (0.35 mg/L); no TN distributions were calculated for High Plains due to lack of data. The 25th percentile of NO_X-N medians was calculated for all level III ecoregions, and ranged from 0.04 mg/L for the Southwest Tablelands to 0.34 mg/L for the Texas Blackland Prairies, a similar range the range observed at the basin level. The High Plains exhibited the highest 25th percentile of median NO_X-N (0.69 mg/L), but only four data points contributed to the frequency distribution for this ecoregion.

The 25th percentiles of the median chl-a concentrations ranged from 3.00 µg/L to 4.30 µg/L for 73% of the level III ecoregions across Texas (Table 1.1.3). Two level III ecoregions had 25th percentile median chl-a concentrations that were greater than 4.30 µg/L with the highest 25th percentile of 9.81 µg/L observed in Central Great Plains. The 25th percentile was not calculated for the High Plains ecoregion, because only two medians were available.

Table 1.1.3. Frequency distribution of median nutrient and chlorophyll-a concentrations for streams and rivers among level III ecoregions in the State of Texas, 2000-2010; these distributions are based on the reduced data with select monitoring type codes excluded.

Total Phosphorus (TP; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	29	0.005	0.050	0.060	0.100	0.358	0.626	0.790
25-High Plains	3	--	--	--	0.230	--	--	--
26-Southwestern Tablelands	44	0.020	0.043	0.060	0.075	0.130	0.449	1.130
27-Central Great Plains	56	0.050	0.060	0.060	0.080	0.216	0.530	2.975
29-Cross Timbers	109	0.029	0.060	0.060	0.082	0.220	0.652	1.960
30-Edwards Plateau	100	0.007	0.020	0.050	0.055	0.060	0.060	1.895
31-Southern Texas Plains	47	0.002	0.050	0.060	0.085	0.132	0.167	0.323
32-Texas Blackland Prairies	207	0.020	0.050	0.060	0.070	0.190	0.865	4.200
33-East Central Texas Plains	74	0.050	0.060	0.096	0.308	0.829	1.140	7.430
34-Western Gulf Coastal Plains	226	0.060	0.100	0.170	0.395	0.984	1.513	3.280
35-South Central Plains	182	0.023	0.060	0.076	0.114	0.170	0.318	3.300

Total Nitrogen (TN; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	24	0.56	0.89	0.99	1.25	2.56	3.16	3.68
25-High Plains	2	3.18	--	--	3.72	--	--	4.25
26-Southwestern Tablelands	16	0.40	0.47	0.60	1.02	2.44	5.42	6.18
27-Central Great Plains	28	0.89	1.05	1.36	1.44	2.20	4.41	11.60
29-Cross Timbers	60	0.47	0.84	1.00	1.22	2.00	5.42	24.95
30-Edwards Plateau	62	0.24	0.30	0.35	0.51	0.90	1.43	6.98
31-Southern Texas Plains	24	0.53	0.82	0.98	1.70	3.06	6.96	16.30
32-Texas Blackland Prairies	139	0.49	0.74	0.97	1.57	3.25	7.35	12.69
33-East Central Texas Plains	54	0.56	0.84	1.62	3.21	8.92	10.53	82.50
34-Western Gulf Coastal Plains	51	0.55	0.94	1.34	1.63	5.14	5.91	8.06
35-South Central Plains	75	0.56	0.66	0.81	0.95	1.17	1.25	9.42

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	28	0.04	0.04	0.08	0.39	0.72	0.97	1.73
25-High Plains	4	0.48	--	0.69	1.69	5.10	--	12.56
26-Southwestern Tablelands	46	0.02	0.02	0.04	0.09	0.67	3.79	6.86
27-Central Great Plains	55	0.02	0.04	0.11	0.23	1.68	3.59	11.30
29-Cross Timbers	105	0.02	0.06	0.10	0.20	0.66	1.75	19.31
30-Edwards Plateau	117	0.02	0.04	0.08	0.21	0.57	1.25	5.36
31-Southern Texas Plains	45	0.02	0.05	0.13	0.28	1.10	4.68	15.95
32-Texas Blackland Prairies	179	0.04	0.15	0.34	0.80	2.15	6.09	11.70
33-East Central Texas Plains	85	0.02	0.10	0.18	0.64	6.21	8.76	70.60
34-Western Gulf Coastal Plains	138	0.02	0.04	0.09	0.33	1.35	3.87	6.72
35-South Central Plains	171	0.04	0.04	0.06	0.15	0.30	0.67	12.39

Phosphate (PO_4-P; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	26	0.006	0.014	0.040	0.060	0.074	0.410	0.570

25-High Plains	3	0.060	--	--	0.060	--	--	1.860
26-Southwestern Tablelands	46	0.020	0.020	0.040	0.040	0.060	0.140	0.780
27-Central Great Plains	56	0.020	0.040	0.040	0.056	0.207	0.343	2.660
29-Cross Timbers	150	0.003	0.007	0.040	0.040	0.060	0.382	1.790
30-Edwards Plateau	100	0.020	0.020	0.020	0.040	0.040	0.060	1.800
31-Southern Texas Plains	41	0.006	0.040	0.040	0.060	0.060	0.110	0.205
32-Texas Blackland Prairies	201	0.010	0.020	0.040	0.040	0.086	0.730	3.860
33-East Central Texas Plains	81	0.020	0.040	0.040	0.075	0.592	0.958	8.010
34-Western Gulf Coastal Plains	232	0.020	0.050	0.080	0.275	0.930	1.666	3.480
35-South Central Plains	173	0.010	0.020	0.040	0.050	0.069	0.197	2.920

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	16	3.00	4.81	8.22	16.8	22.3	31.8	56.7
25-High Plains	2	54.8	--	--	55.3	--	--	55.9
26-Southwestern Tablelands	11	3.00	3.00	4.03	11.5	31.7	38.2	38.6
27-Central Great Plains	19	3.00	5.28	9.81	19.4	33.6	44.7	72.2
29-Cross Timbers	67	3.00	3.29	3.30	7.63	13.7	17.9	39.9
30-Edwards Plateau	42	2.00	3.00	3.00	3.00	5.00	6.66	47.2
31-Southern Texas Plains	22	3.00	3.00	3.00	4.30	8.69	10.4	24.0
32-Texas Blackland Prairies	52	2.00	3.00	3.00	3.52	5.90	13.5	25.2
33-East Central Texas Plains	34	3.00	3.00	3.20	4.44	6.79	12.0	76.8
34-Western Gulf Coastal Plains	50	0.73	3.00	3.72	5.64	9.59	27.8	100
35-South Central Plains	37	3.00	3.00	3.00	4.10	11.4	22.9	41.3

Least Disturbed Streams Frequency Distributions

A subset of streams included in the general stream condition database was identified by TCEQ as least disturbed stream sites, and the frequency distributions of nutrient and chl-a median concentrations calculated from the least disturbed streams dataset are presented in Table 1.1.4. In general, the 75th percentile distribution of the median phosphorus, nitrogen and chl-a concentrations of the least disturbed sites fell within range of the of the 25th percentile median concentrations of the respective parameters calculated from the general streams database considering both basin and level III ecoregion spatial scales. The only exception was the 75th percentile NO_X-N concentration (0.71 mg/L) which exceeded the range of 25th percentile NO_X-N concentrations (0.04-0.69 mg/L NO_X-N) observed in the general stream database across level III ecoregions.

Table 1.1.4. Frequency distribution of median nutrient and chlorophyll-a concentrations among least disturbed stations on Texas streams, 2000-2010; these distributions are based on the reduced data with select monitoring code types excluded.

Parameter	Count	MIN	10th	25th	Median	75th	90th	MAX
Total Phosphorus (TP)	28	0.023	0.057	0.060	0.060	0.145	0.191	1.980
Total Nitrogen (TN)	12	0.47	0.57	0.67	0.97	1.59	7.86	11.60
Nitrate plus Nitrite-Nitrogen (NO_X-N)	26	0.02	0.03	0.04	0.13	0.71	6.41	11.70
Ortho-Phosphate (PO_4-P)	27	0.005	0.020	0.040	0.040	0.055	0.092	2.485
Fluorometric Chlorophyll-a (Chl-a)	13	3.00	3.00	3.00	3.30	3.94	4.80	6.55

USEPA Aggregate Ecoregions

The USEPA has suggested nutrient criteria for TP, TN and chl-a for the 14 aggregate ecoregions in the United States, and the five that partially lie within Texas are presented in Table 1.1 5. The USEPA-suggested nutrient criteria from these five aggregate ecoregions range from 0.01 to 0.13 mg/L of TP (USEPA 2000), and most (90%) of the 25th percentiles of medians calculated from all streams at the level III ecoregion fell within the upper end of this recommended range. However, the 25th percentile of TP for Western Gulf Coastal Plain (0.170 mg/L), as well as, the 75th percentile of least disturbed sites (0.15 mg/L) was greater than USEPA's recommended range. The range of USEPA suggested criteria for total nitrogen among the aggregate ecoregions in Texas was 0.12 to 0.88 mg/L (USEPA 2000) compared to the range of 25th percentile medians across all streams at level III ecoregion calculated in this study which ranged from 0.35 to 1.62 mg/L. 70% of the level III ecoregion 25th percentiles were greater than 0.88 mg/L, and the 75th percentile of the least disturbed streams was almost double the upper limit recommended by the USEPA. The 25th percentile of median chl-a concentrations were also typically greater than the range suggested for the aggregate ecoregions in Texas. The 25th percentile of median chlorophyll-a concentrations from general stream conditions database ranged from 3.00-9.81 µg/L while the range in USEPA suggested criteria was 0.93 to 3.00 µg/L. The 75th percentile of the least disturbed streams (3.94 µg/L) was slightly greater than the USEPA's recommendations. These differences highlight the fact that local and regional impacts can influence the distribution of data, and that criteria specific to an area (i.e., basin or ecoregion or state) should be developed to take into account variations that can occur at spatial scales smaller than the aggregate ecoregion.

The development of frequency distributions from median parameter concentrations is an important first step in the development of nutrient criteria, and the 25th and 75th percentile methods recommended by the USEPA (2000) should be used as a guide when setting criteria for specific geospatial regions. The frequency distribution is also a good method to estimate the number of sites within a spatial scale (e.g., basins, ecoregions) that could exceed the developed criteria. However, this study, as well as others (Ice et al. 2003; Binkley 2004; Longing and Haggard 2010), have shown that the 25th frequency distribution can vary from one basin or ecoregion to another and at different spatial scales. These studies have shown that 25th percentiles based on regional data often significantly differ from that developed for the aggregate ecoregions where USEPA suggested criteria were often more conservative than criteria developed for a specific area (Evans-White, 2013). The frequency distribution method should only be one of many tools used to support the development of numeric nutrient criteria. The Science Advisory Board (SAB) has advised the USEPA that the stressor-response approach is a legitimate, scientifically based method for developing nutrient criteria when correctly applied, and this approach is the focus of the following sections.

Table 1.1.5. USEPA recommended nutrient criteria for streams and rivers for total phosphorus, total nitrogen, and chlorophyll-a for the aggregate ecoregions that in the State of Texas (USEPA, 2000).

Aggregate Ecoregion	Total Phosphorus (mg/L)	Total Nitrogen (mg/L)	Chlorophyll-a (mg/L)
Aggregate Ecoregion II	0.01	0.12	1.08
Aggregate Ecoregion III	0.02	0.38	1.78
Aggregate Ecoregion IV	0.02	0.56	2.40
Aggregate Ecoregion V	0.07	0.88	3.00
Aggregate Ecoregion IX	0.04	0.69	0.93 S[1]
Aggregate Ecoregion X	0.13	0.76	2.10 S[1]

[1]Chlorophyll-a measured by spectrophotometric method with acid correction

1.2. GROUPING STREAM AND RIVER STATIONS WITH SIMILAR NUTRIENT AND BIOLOGICAL CONDITIONS BY THRESHOLDS IN GEOSPATIAL VARIABLES

Methods

We conducted Classification and Regression Tree (CART) analyses on the median database for streams and rivers (described in Section 1.1) to group stream stations by watershed attributes into similar nutrient conditions. The focus (dependent) nutrient variables of these analyses were median TP and TN concentrations (TCEQ parameter codes 00665 and 00600C, respectively), while focus biological variables were median Secchi transparency and chlorophyll-a measured spectrophotometrically and fluorometrically (TCEQ parameter codes, 00078C, 32211, and 70953, respectively). The watershed attributes considered as predictor (independent) variables were divided into 3 categories: 1) land use/land cover, 2) permitted municipal waste water treatment plant (WWTP) flow, and 3) regions. Land use/land cover was further divided into percent developed, percent agriculture, percent developed + agriculture, percent forested, and percent wetland. Permitted municipal waste water treatment plant discharge was used as a predictor variable both unweighted (mgd) and weighted by watershed area (mgd/km^2). Regions considered in the analysis were level III ecoregions, basin, and basin by level III ecoregion.

CART analysis is a means to reduce data based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). For models with numeric predictor variables (i.e. LULC and permitted WWTP discharge, but not regions), the models with the greatest explanatory power were followed by non-parametric changepoint analysis (nCPA) in R.2.9.1 to determine model statistical significance (p_{perm}<0.05) and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median. A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 1.10, the statistical code and raw output generated for each CART and nCPA analysis conducted for this study on geospatial variability in Texas streams and rivers has been compiled.

Results and Discussion

In this section we explored the observed variability in median TP, TN, and chl-a concentrations and Secchi transparencies measured in Texas streams and rivers based on geospatial groupings. We investigated grouping streams and rivers in Texas based on sources of nutrients that have a known impact on water quality such as runoff from adjacent land use and point source discharges. These groupings included watershed land use (e.g., developed, forest, wetland), permitted municipal wastewater treatment plant (WWTP) discharge, and similar geographical regions (i.e., by basin and ecoregion).

Land use/Land cover

Many studies across the US have shown that in-stream nutrient concentrations are related to watershed land use, and nutrients tend to increase with increasing human activity and development within the catchment (Ahearn et al., 2005; Brett et al., 2005; Buck et al., 2004). Conversely, nutrient concentrations are often less in forested catchments. Most studies focused on correlation between nutrients and land use and did not identify landuse thresholds. For Texas streams and rivers, land use/land cover (LULC) categories were weak predictors for TP and TN concentrations (r^2=0.12 and 0.05), although a positive relationship between nutrients and developed land was observed. The %Developed LULC was the strongest predictor among LULC categories for both TP and TN (Fig. 1.3.1A-B). Threshold %Developed LULC was ~10%, indicating that even low level development can affect nutrient concentrations. No statistically significant models based on LULC were identified for chlorophyll-a or Secchi transparency.

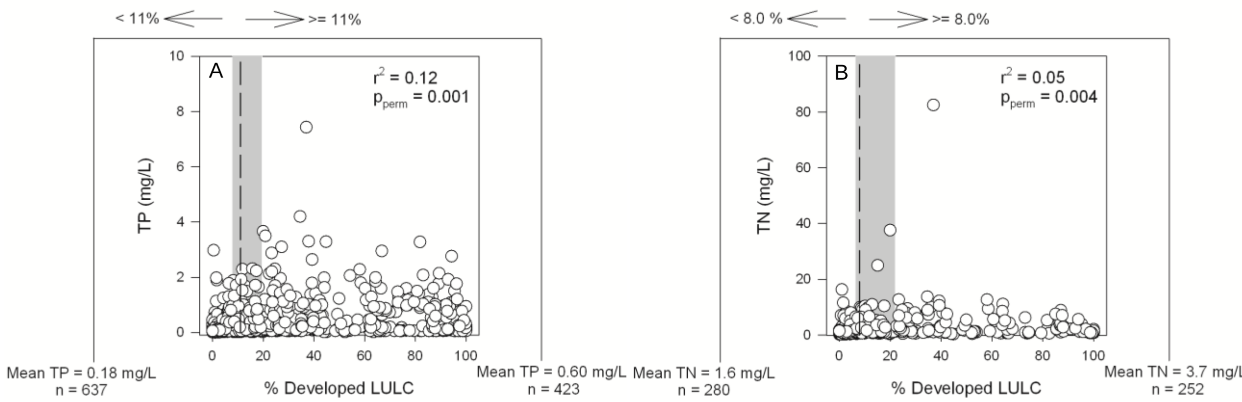


Figure 1.2.1. The relationship between median (A) total phosphorus (TP) and (B) total nitrogen (TN) and percent developed land use (% Developed LULC) across Texas streams and rivers. Thresholds are based on classification and regression tree analysis (CART).

Permitted Municipal Waste Water Treatment Plant Flow

Permitted municipal waste water treatment plant discharge generally increases nutrient concentrations, particularly TP concentrations, in receiving streams. The effects of WWTP discharge can often be observed for several kilometers downstream (Haggard et al. 2010), especially under base flow conditions. The extent of nutrient concentration increase is dependent upon the volume of effluent discharge by the plant, the permitting levels of the WWTP, and the size and background flow regime of the receiving stream. Because information on permitting levels and discharge volumes for Texas streams and rivers was not available, we divided permitted municipal WWTP flow by the catchment area to approximate the proportion of in-stream base flow contributed by WWTP discharge. Unweighted permitted WWTP discharge was a moderate predictor for TP and a weak predictor for TN concentrations (r^2=0.17 and 0.04, respectively; Fig. 1.2.2A-B). For both TP and TN concentration, a threshold concentration of approximately 4 mgd was identified above which TP and TN concentrations were approximately 4x and 2x greater on average, respectively. Weighting WWTP discharge by watershed area (mgd/km^2) resulted in near doubling of CART model explanatory power for both TP and TN (r^2=0.32 and 0.09, respectively; Fig. 1.2.3A-B). The area-weighted WWTP threshold for both TP and TN concentrations was 0.031 mgd/km^2. Above this threshold, TP and TN concentrations were approximately 5x and 2.5x greater on average. No statistically significant geospatial models based on WWTP flow were identified for chlorophyll-a or Secchi transparency. The calculated parameter area-weighted permitted municipal WWTP discharge has been provided to TCEQ in the streams and rivers median water quality database.

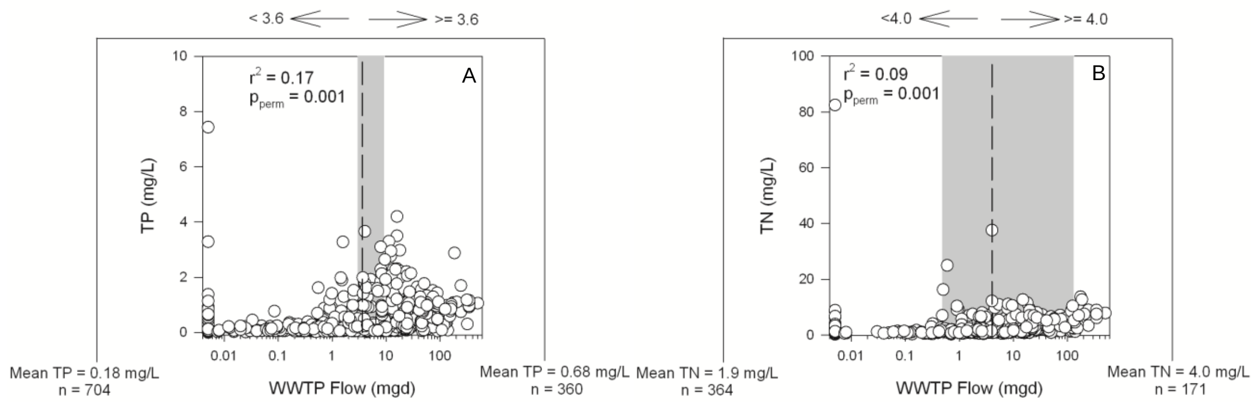


Figure 1.2.2. The relationship between median (A) total phosphorus (TP) and (b) total nitrogen (TN) and permitted waste water treatment plant (WWTP) flow before weighting by watershed area across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

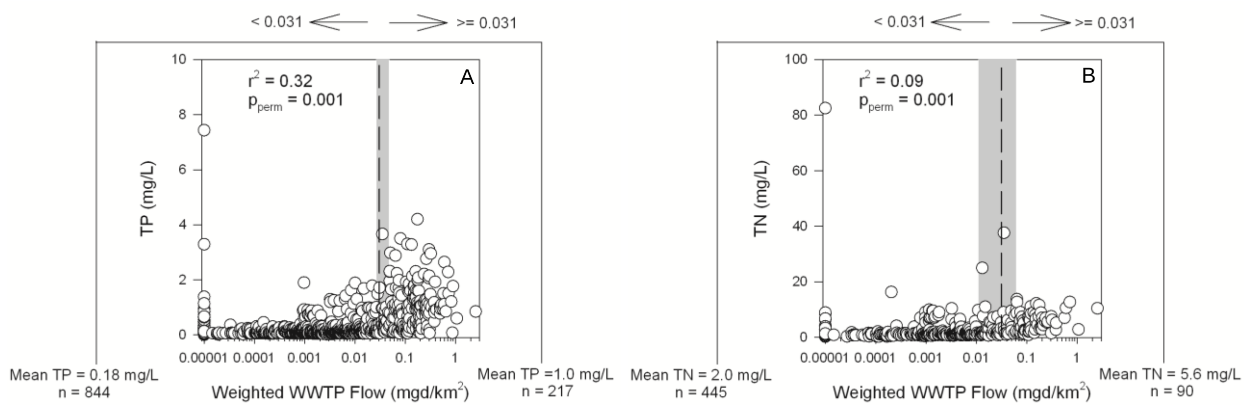


Figure 1.2.3. The relationship between median total phosphorus (TP) and permitted waste water treatment plant (WWTP) flow weighted by watershed area across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

Regions

The EPA's original nutrient criteria were set based on regional groupings, but at the coarse spatial scale of the aggregate nutrient ecoregion (USEPA, 2000). Here, we applied the same principle of grouping streams and rivers based on regional similarity, but on a smaller scale. We investigated grouping streams based on basin, level III ecoregion, and a combination of basin and level III ecoregion (basin-ecoregion III). The basin-ecoregion III parameter was the most refined scale that we analyzed and was the strongest regional predictor for both TP and TN. CART groupings of "low" and "high" TP and TN stations are summarized in

Table 1.2.1. For both TP and TN, CART separated out less than 20% of basin-ecoregion III combinations and a maximum of 21% of stations as “high” nutrient regions. For both nutrient variables, median TN and TP concentrations were approximately 4x greater in the high vs. low nutrient groups.

For biological variables, CART models based on regional groupings had low explanatory power (r^2<0.10), except for fluorometric chl-a (r^2=0.15). Groupings of “low” and “high” chlorophyll-a and Secchi transparency are summarized in Table 1.2.2. As with TP and TN concentrations, for fluorometric chl-a CART separated out less than 20% of basin-ecoregion II combinations and only approximately 10% of stations as having “high” chl-a concentrations, but median chl-a concentrations were approximately 3x higher on average for these stations.

Table 1.2.1 Groupings of Basin-Ecoregion III areas by “low” and “high” median total phosphorus (TP) and total nitrogen (TN) concentrations in Texas streams and rivers based on classification and regression tree analysis (CART).

Median TP			Median TN		
“Low”, n = 884 Mean TP= 0.21 mg/L		“High”, n = 193 Mean TP= 0.96 mg/L	“Low”, n = 469 Mean TN = 1.9 mg/L		“High”, n = 66 Mean TN = 7.2 mg/L
1-26	12-32	2-32	1-26	12-30	8-33
2-26	12-34	3-33	2-26	12-32	10-34
2-27	13-34	8-33	2-27	12-34	12-33
2-29	14-26	10-34	2-29	13-34	19-33
2-33	14-27	12-27	2-35	14-26	19-34
2-35	14-29	12-33	3-32	14-27	21-31
3-32	14-30	14-25	3-33	14-29	22-34
3-35	14-32	19-33	3-35	14-30	8-33
4-35	14-33	20-34	4-35	14-32	
5-32	14-34	22-34	5-32	18-30	
5-34	18-30		5-34	18-32	
5-35	18-33		5-35	18-33	
6-35	18-34		6-35	13-34	
7-34	19-30		7-34	19-30	
8-29	19-32		8-29	19-31	
8-32	19-34		8-32	21-30	
8-35	21-30		8-35	21-33	
9-34	21-31		9-34	23-24	
10-35	21-33		10-35	23-30	
11-34	21-34		11-34	23-31	
12-25	23-31		12-25	23-34	
12-26	23-34		12-26		
12-29	24-34		12-27		
12-30			12-29		

Table 1.2.2. Groupings of Basin-Ecoregion III areas by “low” and “high” median chlorophyll-a concentrations measured spectrophotometrically (Chl-a Spec) and fluorometrically (Chl-a Fluoro) and Secchi transparency in Texas streams and rivers based on classification and regression tree analysis (CART).

Median Chl-a Spec $r^2 = 0.07$		Median Chl-a Fluoro $r^2 = 0.15$		Median Secchi $r^2 = 0.04$	
“Low”, n= 427 Mean Chl-a = 7.6	“High”, n= 170 Mean Chl-a = 12	“Low” n= 319 Mean Chl-a = 8.7	“High” n= 33 Mean Chl-a = 26	“Low” 577 Mean Secchi= 0.46	“High” n= 400 Mean Secchi= 0.60
1-26	2-26	1-26	2-26	2-26	1-26
2-27	2-33	2-27	4-35	2-29	2-27
2-29	3-33	2-35	14-34	2-32	2-33
2-32	4-33	3-32	20-34	3-33	2-35
2-33	4-35	3-33	21-34	3-35	3-32
2-35	5-33	3-35	23-34	4-33	5-35
3-32	6-35	4-33	24-34	4-35	8-33
3-35	7-34	5-33		5-32	8-34
5-32	8-33	5-34		5-33	8-35
5-34	8-34	5-35		5-34	10-34
5-35	10-34	6-35		6-33	10-35
8-29	14-33	7-34		6-35	11-34
8-32	18=-32	8-29		7-34	12-25
8-35	18-33	8-32		8-29	12-27
10-32	18-34	8-33		8-32	12-33
10-35	19-31	8-34		9-34	13-34
11-34	19-32	8-35		12-26	14-29
12-26	19-33	9-34		12-29	14-30
12-27	20-34	10-34		12-30	18-30
12-29	21-30	10-35		12-32	18-33
12-30	21-34	11-34		12-34	20-33
12-32	22-34	12-26		14-26	21-31
12-33	23-30	12-27		14-27	21-33
12-34	23-34	12-29		14-32	23-24
13-34	24-34	12-30		14-33	24-34
14-25		12-32		14-34	
14-26		12-33		1832	
14-27		12-34		18-34	
14-29		13-34		19-30	
14-30		14-27		19-31	
14-32		14-29		19-32	
14-34		14-30		19-33	
18-30		14-32		19-34	
19-30		14-33		20-34	
19-34		18-30		21-30	
20-33		18-32		21-34	
21-31		18-33		22-34	
21-33		19-30		23-30	
23-24		19-32		23-34	
23-31		19-33			
		21-30			
		21-31			
		22-34			
		23-24			
		23-30			
		23-31			

Geospatial analysis summary

A summary of the strongest CART models for median TP, TN, and chl-a concentrations and Secchi transparency is available in Table 1.2.3. In general, geospatial grouping schemes were more effective for nutrient concentrations than for biological variables in Texas streams and rivers. In turn, geospatial groupings were more effective for TP concentrations than TN. Model strength was consistently lower for TN than for TP, but many of the TN analyses indicated identical or similar thresholds in geospatial variables and regional groupings to those identified for TP. The models with the greatest predictive power for grouping nutrients were TP vs. weighted WWTP flow, TP vs. Basin-Ecoregion III, and TN vs. Basin-Ecoregion III.

Table 1.2.3. Summary of classification and regression tree (CART) models with the greatest statistical power in each geospatial category including land use/ land cover (LULC), waste water treatment plant (WWTP) flow, and region across streams and rivers in Texas. NS=not significant; Groups=no numerical thresholds.

Parameter	Geospatial Category	Predictor	Threshold	Model r^2
Total Phosphorus	LULC	% Developed	11%	0.12
	WWTP Flow	Weighted	0.031 mgd/km^2	0.32
	Region	Basin-Ecoregion III	Groups	0.24
Total Nitrogen	LULC	% Developed	8%	0.05
	WWTP Flow	Weighted by watershed area	0.031 mgd/km^2	0.09
	Region	Basin-Ecoregion III	Groups	0.14
Chl-a Spectrophotometric	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Region	Basin-Ecoregion III	Groups	0.07
Chl-a Fluorometric	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Region	Basin-Ecoregion III	Groups	0.15
Secchi Transparency	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Region	Basin-Ecoregion III	Groups	0.04

1.3. STRESSOR-RESPONSE ANALYSIS ON THE STREAMS AND RIVERS MEDIAN WATER QUALITY DATABASE

Methods

We conducted CART analyses on the median database for streams and rivers (see Section 1.1) to identify thresholds in nutrient concentrations that resulted in measurable changes in common biological responses. The biological (dependent) variables included in the analyses were: median Secchi depth (m; parameter code 00078C), median 24 hour dissolved oxygen (DO) flux (parameter code 89856C), median chlorophyll-a chl-a measured with spectrophotometry (chl-a spec; parameter code 32211), and median chlorophhyll-a measured with fluorometry (chl-a fluoro; parameter code 70953). The nutrient (independent) variables included in the analysis were total phosphorus (TP; 00665), total nitrogen (TN;

00600C), nitrite plus nitrate-nitrogen (NO_x-N; 00631C), ammonia-nitrogen (NO_4-N; 00610), and soluble reactive phosphate (SRP; 00671C).

CART analysis is a form of data reduction that aims to: 1) quantify thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables, and 2) identify hierarchical structure in independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian et al. 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). The models with the greatest explanatory power were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance ($p_{perm}<0.05$) and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value.

We first ran CART models using all median data from the streams and rivers median database. Secondarily, we ran CART models after limiting data to stations receiving low or high permitted municipal WWTP discharge weighted by area (stations receiving < 0.031 or ≥ 0.031 mgd/km^2, respectively; see Section 1.1 for details). These datasets are subsequently referred to as "all," "low," and "high" weighted WWTP flow datasets subsequently. Because CART analysis involves recursive partitioning, models may sometimes be over-fit (i.e. too many independent variables that decrease the statistical rigor of final model). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model crossvalidations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 1.11, the statistical code and raw output generated for each CART and nCPA analysis conducted on stressor-response relationships in Texas streams and rivers has been compiled.

Results and Discussion

Secchi Transparency

In aquatic systems, Secchi transparency is a proxy for water clarity and often correlates with nutrient concentrations, decreasing as nutrients increase. In Texas rivers and streams, TP was the strongest predictor for Secchi transparency, and model strength for TP was ~4x greater than for any other potential nutrient stressor. TP thresholds ranged from 0.063 – 0.091 mg/L in the three WWTP flow-based datasets. Identical TP thresholds for Secchi transparency response were identified in the “all” and “low” datasets (0.063 mg/L; Fig. 1.3.1A-B), indicating that stations that were potentially strongly affected by WWTP flow were not driving the CART analysis for the “all” dataset. A slightly higher TP threshold was identified in the “high” dataset (0.091 mg/L; Fig. 1.3.1C). However, the analysis was potentially constrained by the number of medians required to form a split, i.e. no fewer than 10 medians were allowable to the right or left of a split. If this number were reduced, it is possible that a lower TP threshold would be identified.

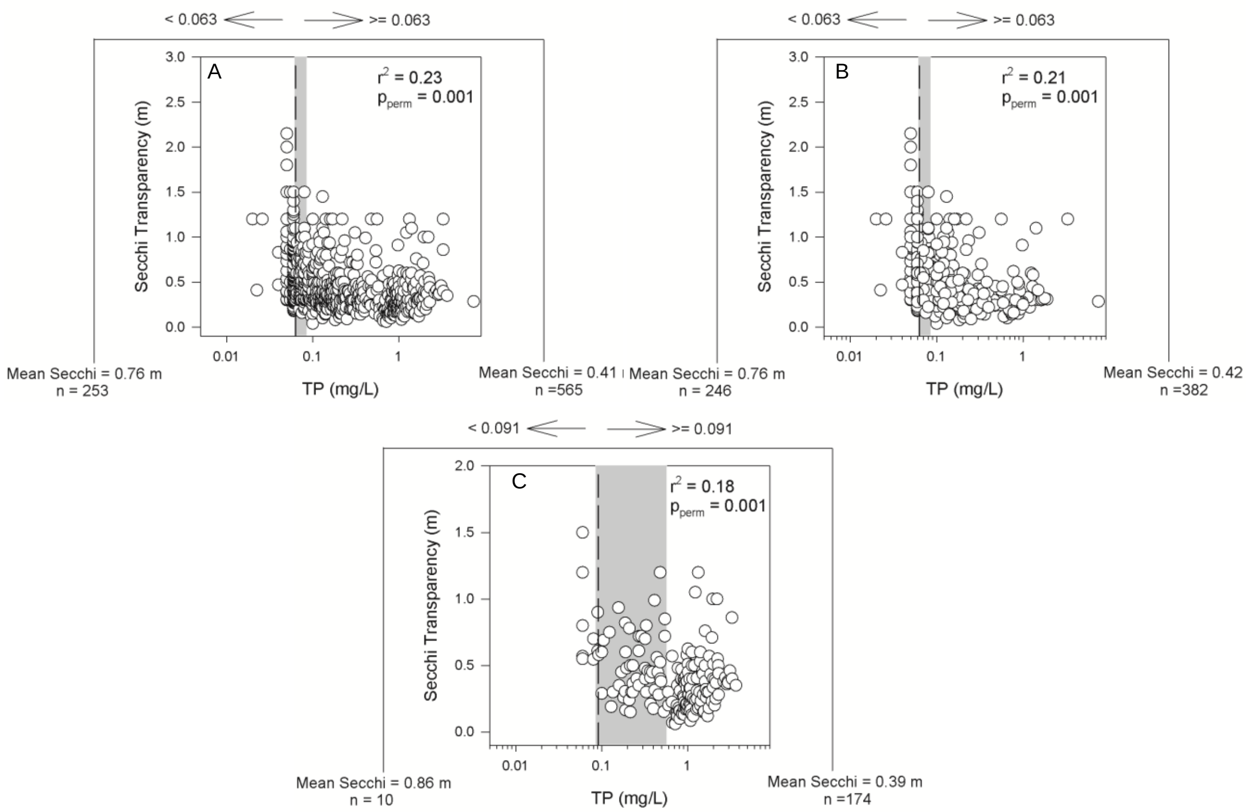


Figure 1.3.1. The relationship between Secchi transparency and total phosphorus (TP) for the all medians (A), low area-weighted wastewater treatment plant (WWTP) flow (B), and high area-weighted WWTP flow (C) geospatial data subsets across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

A weak relationship was found between Secchi transparency and TN concentration in the "all" and "low" area-weighted permitted WWTP flow datasets. As was observed for TP concentrations, the TN threshold identified in both datasets was identical (0.70 mg/L; Fig. 1.3.2 A-B); therefore, the stations that were potentially most strongly influenced by WWTP were not driving the CART analysis on Secchi transparency vs. TN concentration in the "all" weighted flow dataset. No statistically significant TN threshold for Secchi transparency response was found in the "high" weighted flow dataset (Fig. 1.3.2C). Most of the median TN concentrations at stations in the "high" weighted flow dataset exceeded the threshold of 0.70 mg/L identified in the "all" and "low" weighted flow datasets, suggesting that TN concentrations are often elevated above potentially limiting conditions in streams that are highly impacted by permitted municipal waste water discharge.

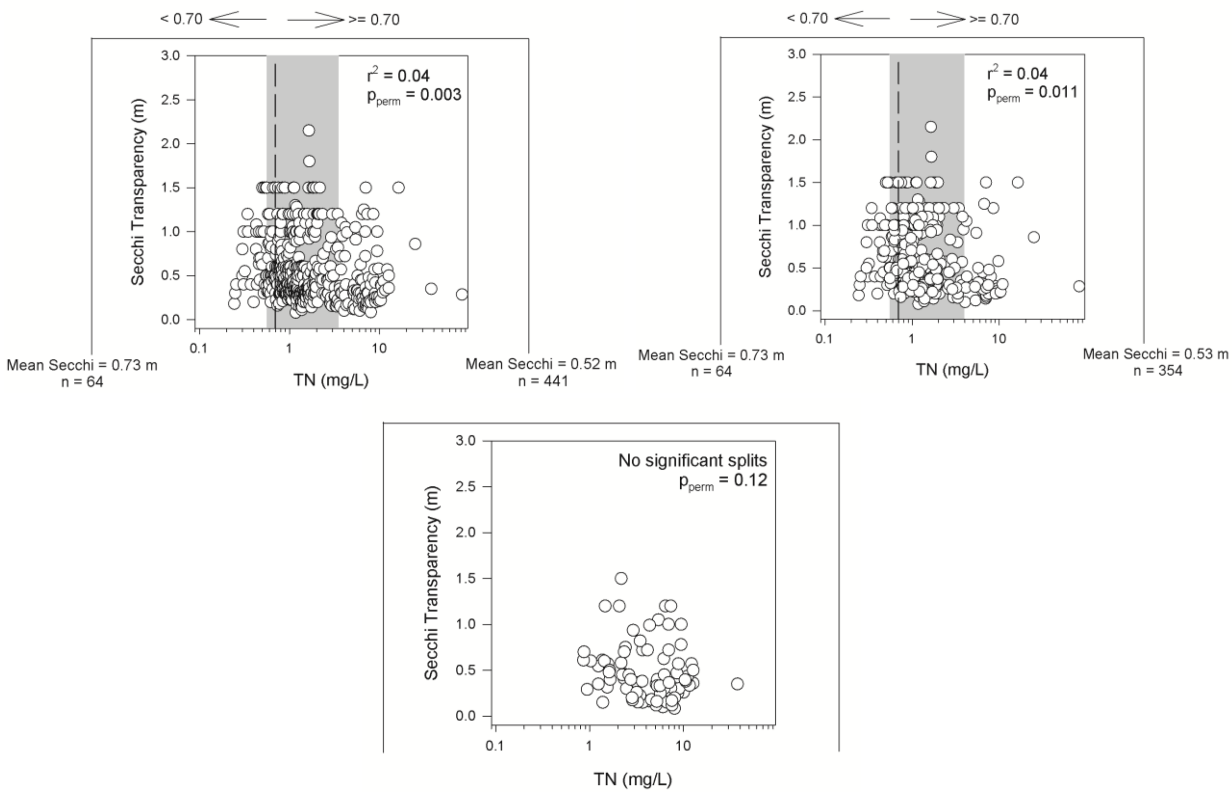


Figure 1.3.2. The relationship between Secchi transparency and total nitrogen (TN) for the all medians (A), low area-weighted wastewater treatment plant (WWTP) flow (B), and high WWTP area-weighted flow (C) geospatial data subsets across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

Discussion in the scientific literature of nutrient criteria based on Secchi transparency is limited for streams and rivers. It is known, however, that suspended algal concentrations are generally positively related with total nutrient concentrations in streams and rivers (Haggard et al., 2013). Because increased sestonic algal turbidity due to nutrient enrichment would play a major part in reduced water clarity, Secchi transparency should hypothetically be negatively, or inversely related to total nutrient concentrations. This is indeed the pattern that we observed in Texas rivers and streams. Furthermore, the TP and TN thresholds for Secchi transparency identified in Texas streams and rivers (i.e., 0.063-0.091 mg TP/L and 0.70 mg TN/L) were within the range of other thresholds that have been identified for suspended algae in the US (e.g., 0.007-2.8 mg TP/L and 0.21-18.7 mg TN/L, Royer et al., 2008). Using Secchi transparency as a response variable for stressor-response analyses may also have the advantage of incorporating the effects on water quality of inorganic turbidity. This could explain why Secchi transparency was more strongly related to TP than TN concentration, as P is more strongly associated with inorganic particles. However, this hypothesis requires further investigation.

Sestonic Chlorophyll-a

In streams and rivers, chl-a and nutrient concentrations are generally positively related (Dodds et al. 2002, 2006, Robertson et al., 2006, Stevenson et al., 2008, Haggard et al. 2013). In Texas streams and rivers, TP was consistently the strongest predictor of both sestonic chl-a, but these models had lower explanatory power than models relating TP and Secchi transparency (r^2 <0.10 vs. r^2~0.20). The TP thresholds for sestonic chl-a were also higher than those identified for Secchi transparency. For chl-a spec (Figs. 1.3.3A,C,E) and chl-a fluoro (Figs. 1.3.3B,D,F), TP thresholds were similar, but not identical, between the "all" and "low" WWTP flow-based datasets. For chl-a spec, TP thresholds were higher than those identified for chl-a fluoro (0.1-0.11 mg/L vs. 0.069-0.079 mg/L). For the "high" area-weighted permitted municipal WWTP flow subset, no statistically significant TP thresholds for either chl-a method were identified.

For chl-a measured spectrophotometrically, TN thresholds were identified in CART for the "all" and "low" area-weighted WWTP flow datasets (Figs. 1.3.3A,C). CART models relating chl-a spec to TN had low explanatory power (r^2=0.04-0.06). The TN thresholds identified for chl-a spec were identical between these two datasets, indicating that stations potentially highly impacted by municipal WWTP flow were not driving the analyses of the "all" area-weighted WWTP flow dataset for models relating chl-a spec and TN, which was consistent with results from these two datasets for all other stressor-response pairings. For chl-a spec, TN thresholds occurred at higher concentrations than for Secchi transparency (1.1. mg/L for chl-a spec vs. 0.70 mg/L for Secchi transparency). No statistically significant TN thresholds were identified for chl-a spec in the "high" area-weighted WWTP flow dataset (fig. 1.3.3.E) or chlorophyll-a measured fluorometrically in any of the area-weighted flow datasets (Figs. 1.3.3B,D,F).

The majority of thresholds based on the relationship between chl-a and nutrient concentrations reported in the literature are based on benthic algal biomass metrics, rather than sestonic chl-a. Suggested criteria based on a variety of methods including regression, regression tree, nCPA and two-dimensional Kologrov Smironov test range from 0.0127-0.043 mg/L for TP and 0.435-0.918mg/L for TN (Dodds et al., 2002, 2006, Robertson et al., 2006, Stevenson et al., 2008). The TN thresholds identified for Texas streams and rivers fall within this range, but the TP thresholds consistently exceed this range.

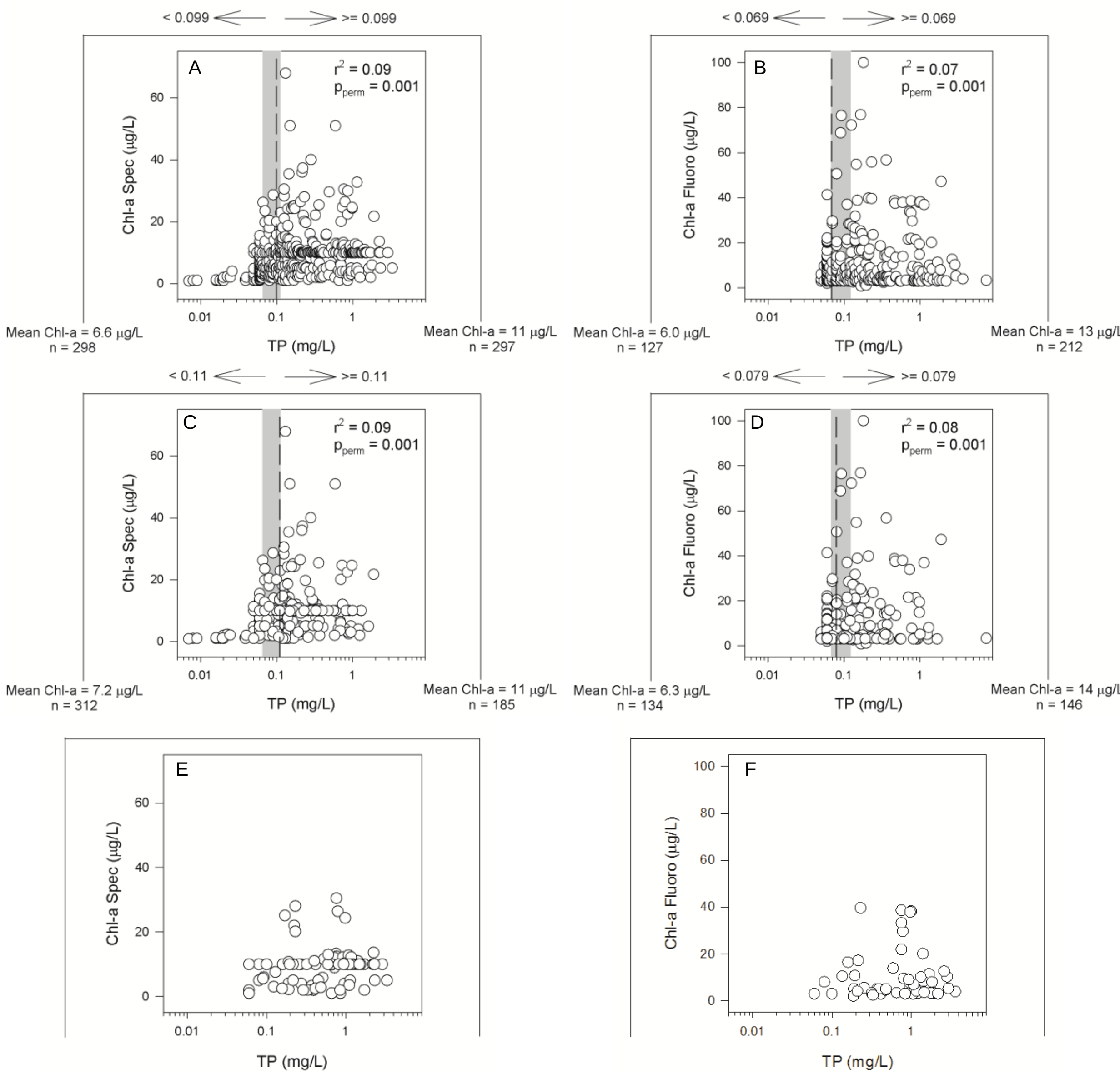


Figure 1.3.3. The relationship between chlorophyll-a (chl-a) measured spectrophotometrically and fluorometrically and total phosphorus (TP) for the all medians (A-B), low area-weighted wastewater treatment plant (WWTP) flow (C-D), and high area-weighted WWTP flow (D-F) geospatial data subsets across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

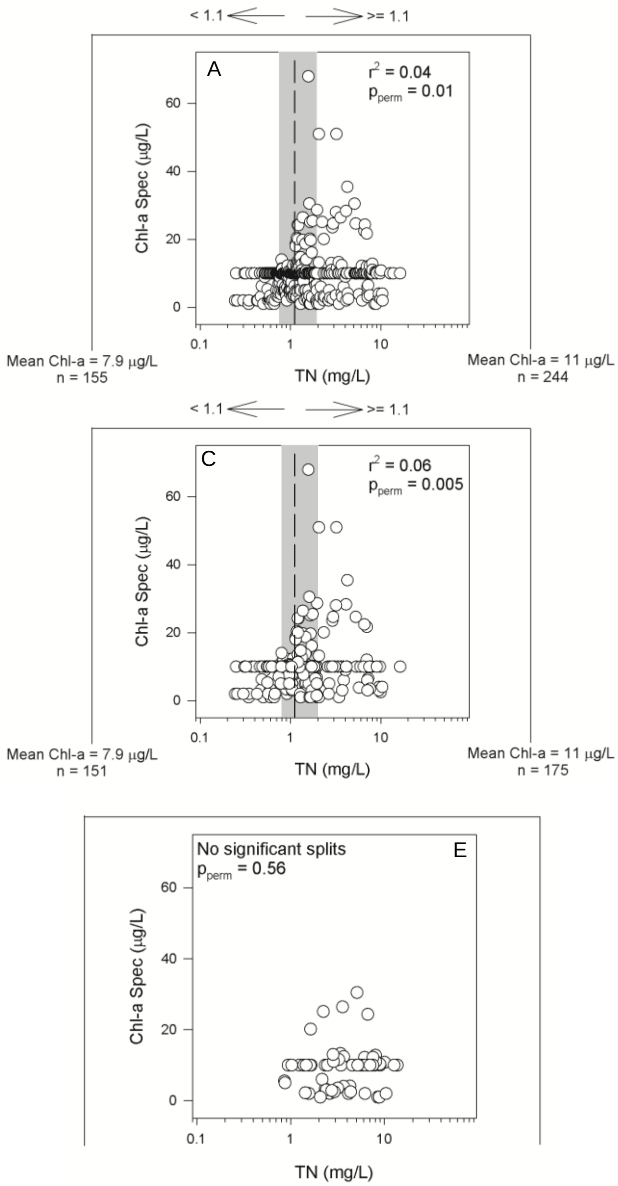


Figure 1.3.4. The relationship between chlorophyll-a (chl-a) measured spectrophotometrically and fluorometrically and total nitrogen (TN) for the all medians (A-B), low area-weighted wastewater treatment plant (WWTP) flow (C-D), and high area-weighted WWTP flow (D-F) geospatial data subsets across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

Table 1.3.1. Thresholds identified in classification and regression tree (CART) analyses of the streams and rivers median water quality database. A dash indicates that no statistically significant threshold could be identified for the given stressor-response pairing. For these pairs, r^2 and p values are italicized. For the high area-weighted WWTP plant flow datasets, CART analysis did not identify models for DO Flux response to any of the nutrient predictor variables for subsequent testing in nCPA. Therefore, r^2 and p values were not available for these stressor-response pairings.

	Biological Variable	TP (mg/L)			TN (mg/L)			NOx-N (mg/L)			NH_4-N (mg/L)			SRP (mg/L)		
		CP	r^2	p	CP	r^2	p	CP	r^2	p	CP	r^2	p	CP	r^2	p
All Stations	Secchi	0.063↓	0.23	0.001	0.70↓	0.04	0.003	--	0.01	0.068	0.051↓	0.06	0.001	0.047↓	0.07	0.001
	DO Flux	--	0.06	0.21	--	0.11	0.13	0.055↓	0.14	0.038	--	0.02	0.68	--	0.02	0.80
	Chl-a Spec	0.099↑	0.09	0.001	1.1↑	0.04	0.010	--	0.02	0.15	0.050↑	0.04	0.001	0.035↑	0.04	0.005
	Chl-a Fluoro	0.069↑	0.07	0.001	--	0.05	0.059	--	0.02	0.28	0.049	0.03	0.030	--	0.02	0.13
Low Flow	Secchi	0.063↓	0.21	0.001	0.70↓	0.04	0.011	0.45↑	0.03	0.005	0.051↓	0.04	0.001	0.047↓	0.05	0.001
	DO Flux	--	0.09	0.097	--	0.12	0.14	0.055↓	0.18	0.021	--	0.06	0.19	--	0.04	0.37
	Chl-a Spec	0.11↑	0.09	0.001	1.1↑	0.06	0.005	--	0.02	0.13	0.049↑	0.03	0.016	0.035↑	0.05	0.006
	Chl-a Fluoro	0.079↑	0.08	0.001	--	0.04	0.084	--	0.02	0.28	0.049	0.03	0.051	--	0.02	0.13
High Flow	Secchi	0.091↓	0.18	0.001	--	0.08	0.12	--	0.02	0.60	0.061↓	0.07	0.019	0.050↓	0.18	0.001
	DO Flux	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Chl-a Spec	--	0.05	0.31	--	0.04	0.56	--	0.09	0.056	0.050	0.15	0.005	--	0.03	0.51
	Chl-a Fluoro	--	0.05	0.49	--	0.05	0.48	--	0.07	0.34	0.15	0.20	0.011	--	0.08	0.20

↓ The value of the response variable decreases with increasing predictor variable values, and vice versa

↑ The value of the response variable increases with increasing predictor variable values, and vice versa

Streams and Rivers Median Stressor Response Summary

For Texas streams and rivers, TP was consistently the strongest nutrient predictor for Secchi transparency and chl-a in the “all”and “low” area-weighted WWTP flow datasets. In contrast, TN was consistently a weak nutrient predictor for Secchi transparency and chl-a in the “all” and “low” datasets. In most cases, statistically significant total nutrient thresholds were not identified for the “high” area-weighted WWTP flow dataset. The TP and TN thresholds identified for biological response ranged from 0.063-0.11 mg/L and 0.70-1.1 mg/L, respectively, across the 3 datasets. Nutrient thresholds identified for Secchi transparency were consistly lower than those identified for chl-a, particularly for chl-a spec. No statistically significant relationships between DO Flux and total nutrients were found in the all station and low flow datasets, and insufficient DO Flux data were available in the high flow dataset for CART analysis. CART models based on total nutrients, especially TP, declined in strength and thresholds increased in magnitude in the high flow dataset in comparison with the all stations and low flow datasets. See Table 1.3.1 for a full summary of results, including models with dissolved nutrient fractions as stressor variables.

Further investigation

DO Flux was a function of chl-a. CART analysis indicated that nutrients were not related to DO Flux in Texas streams and rivers. However, chl-a was a strong predictor for DO Flux in both the “all” stations (Fig. 1.3.5) and “low” area-weighted WWTP flow datasets. Therefore, DO Flux was directly related to algal activity in Texas streams and rivers, a finding that is supported by ecological theory. The less direct link between nutrients and DO Flux may not be apparent in these analyses because nutrient availability is just one of many factors that regulates algal productivity. Light availability is a key physical factor regulating algal productivity that can vary with stream order and riparian LULC, but was not considered in these analyses.

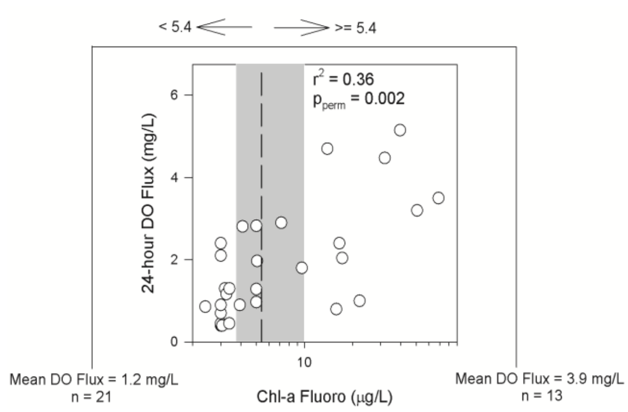


Figure 1.3.5. The relationship between 24-hour dissolved oxygen (DO) flux and chlorophyll-a (Chl-a) measured fluorometrically for the all medians data set across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

Dissolved nutrients important in high flow dataset. CART analysis indicated that models showing the relationship between biological response variables and total nutrient fractions, particularly TP, declined in strength or did not exist in the high flow dataset. However, models of the relationship between biological response variables and dissolved inorganic nutrient fractions increased in strength. For both chl-a spec and chl-a fluoro, the strongest predictor variable was NH_4-N, though the value of the threshold differed between methods (Figs 1.6.5A-B). This finding likely reflects the fact that TP is elevated above threshold levels for many of the high flow streams, resulting in nutrient limitation shifting to other nutrient fractions, in this case, NH_4-N. Aso, the CART model of Secchi transparency vs. SRP (Fig. 1.3.7) was equal in strength to the model of Secchi transparency vs. TP (Fig. 1.3.1). This finding likely reflects that dissolved inorganic P as a proportion of TP is elevated in streams strongly affected by WWTP discharge.

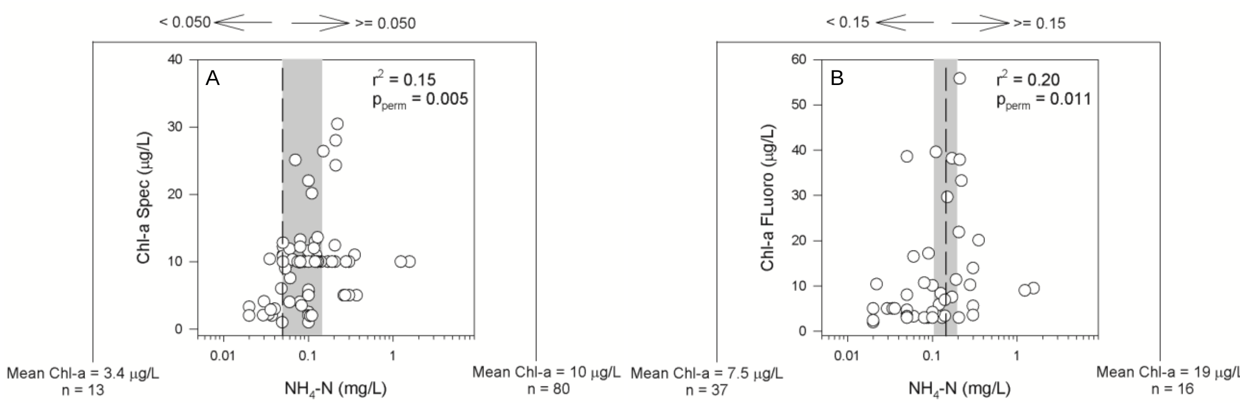


Figure 1.3.6. The relationship between chlorophyll-a measured spectrophotometrically (chl-a spec; A) and chlorophyll-a measured fluorometrically (chl-a fluoro; B) for the high area-weighted wastewater treatment plant (WWTP) dataset across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

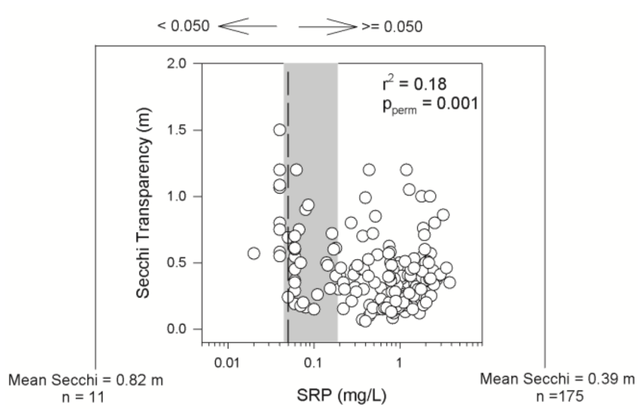


Figure 1.3.7. The relationship between Secchi transparency and soluble reactive phosphorus (SRP) in the high area-weighted wastewater treatment plant (WWTP) flow dataset across Texas streams and rivers showing thresholds based on classification and regression tree analysis (CART).

1.4: STRESSOR-RESPONSE ANALYSIS ON THE STREAMS AND RIVERS BIOASSESSMENT DATABASE

Methods

The Texas Commission on Environmental Quality (TCEQ) provided bioassessment data including Fish Index of Biotic Integrity (IBI), Macrobenthic Rapid Bioassessment Index of Biotic Integrity (RBIBI), aquatic life use (ALU) scores, habitat quality index (HQI) and others from 173 stations that spanned 16 basins. This database was provided in a useable format where Station ID and parameter labels and associated data were in a unique column. Water quality data from the streams and rivers median water quality database were paired with bioassessment data based on Station ID. There were 11 stations in the bioassessment database that did not match Station IDs in the water quality database, so water quality stations located upstream of the missing station on the same or a nearby reach were paired with nine of these bioassessment data points. However, two stations did not have a similar station nearby, and the stations were removed from the bioassessment database.

Median values of each parameter were calculated for various time periods including the period of record (time frame during which all bioassessment data were collected at a given station), year, index period (March 15-October 15), critical period (July 1-September 30), and monthly. Raw water quality and bioassessment data points were also paired in a "matching" dataset. The calculated medians for each Station ID were then compiled into a database for each of the identified time periods. Medians were calculated for various time periods because raw data observations were not consistently paired for some stations, because interannual and seasonal variability can occur in biological and nutrient data within a stream. Stressor-response analyses were subsequently conducted with each of these temporal datasets in order to determine if the time frame over which measures of central tendency were calculated would affect analyses.

We conducted CART analyses on the bioassessment median database for streams and rivers to identify thresholds in nutrient concentrations that resulted in measurable changes in Fish IBI and RBBI. The Habitat Quality Index (HQI) was also included as a potential predictor in these CART models because these biological indices are designed to capture the effect of degraded habitat on biological health. We also constructed CART models to identify nutrient thresholds in some common biological variables that were the primary focus of the previous chapter: median Secchi depth (m; parameter code 00078C), median 24 hour dissolved oxygen (DO) flux (parameter code 89856C), median sestonic chl-a measured with spectrophotometry (chl-a spectro; parameter code 32211)), median chl-a measured with fluorometry (chl-a fluoro; parameter code 70953). The independent variables included in the analysis were total phosphorus (TP; parameter code 00665), total nitrogen (TN parameter code 00600C), nitrite plus nitrate-nitrogen (NO_X-N; parameter code 00631C), and the HQI.

CART analysis is a form of data reduction that aims to: 1) quantify thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables, and 2) identify hierarchical structure in independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). The models with the greatest explanatory power were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance ($p<0.05$) and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value. CART models on the complete dataset were also developed for median Secchi depth (m), median 24 hour DO flux, median chl-a spec, and median chl-a fluoro in order to compare these results to similar models built using the complete stream and river median database (See section 1.3). Because CART analysis involves recursive partitioning, models may sometimes be over-fit (i.e., too many independent variables that decrease the statistical rigor of final model). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 1.12, the statistical code and raw output generated for each CART and nCPA analysis conducted for this study on stressor-response relationships in the Texas streams and rivers bioassessment database has been compiled.

Results and Discussion

Fish IBI

Both nutrients and habitat quality influence fish populations in streams and rivers. Increased nutrient concentrations often negatively impact fish communities, while high habitat quality promotes abundance and diversity. In Texas streams and rivers, we observed these trends, but also variability in thresholds between the multiple time frames for which bioassessment, nutrient, and habitat data were analyzed for central tendency and paired. Fish IBI consistently increased as HQI increased, and thresholds ranged from 18 to 21 (Fig. 1.4.1). HQI was the strongest predictor for Fish IBI when considering the critical period and paired observations datasets (r^2= 0.08 and 0.06). The literature highlights the importance of habitat for fish populations (e.g., Gorman and Karr, 1978), the discussion of thresholds in habitat quality is limited. Thresholds were in range with habitat thresholds for other biological indicators (RBIBI, Secchi and chl-a).

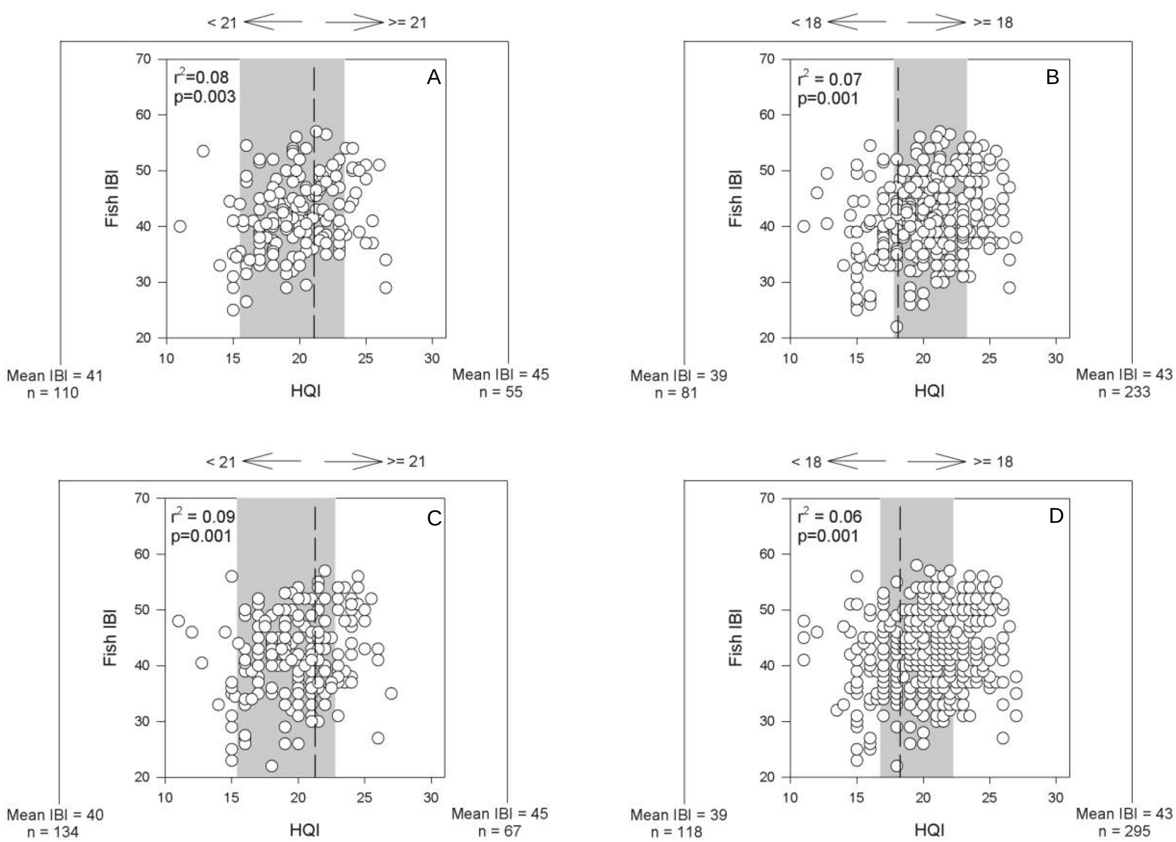


Figure 1.4.1. The relationship between Fish Index of Biotic Integrity (IBI) and habitat quality index (HQI) for Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

For medians calculated for all time periods, Fish IBI decreased as TP increased, and TP thresholds ranged from 0.084-0.098 mg/L (r^2= 0.07-0.13; fig. 1.4.2 A-C). These thresholds fall within the range (0.07-0.139 mg/L) reported by Wang et al., 2007, and Weigel and Robertson, 2007 for other biological variables. No statistically significant TP threshold was identified for fish response in the dataset containing paired observations. In general, when Fish IBI was compared to TP concentration, model strength and statistical probability declined with increasingly narrow time frame over which data were analyzed for central tendency and were paired.

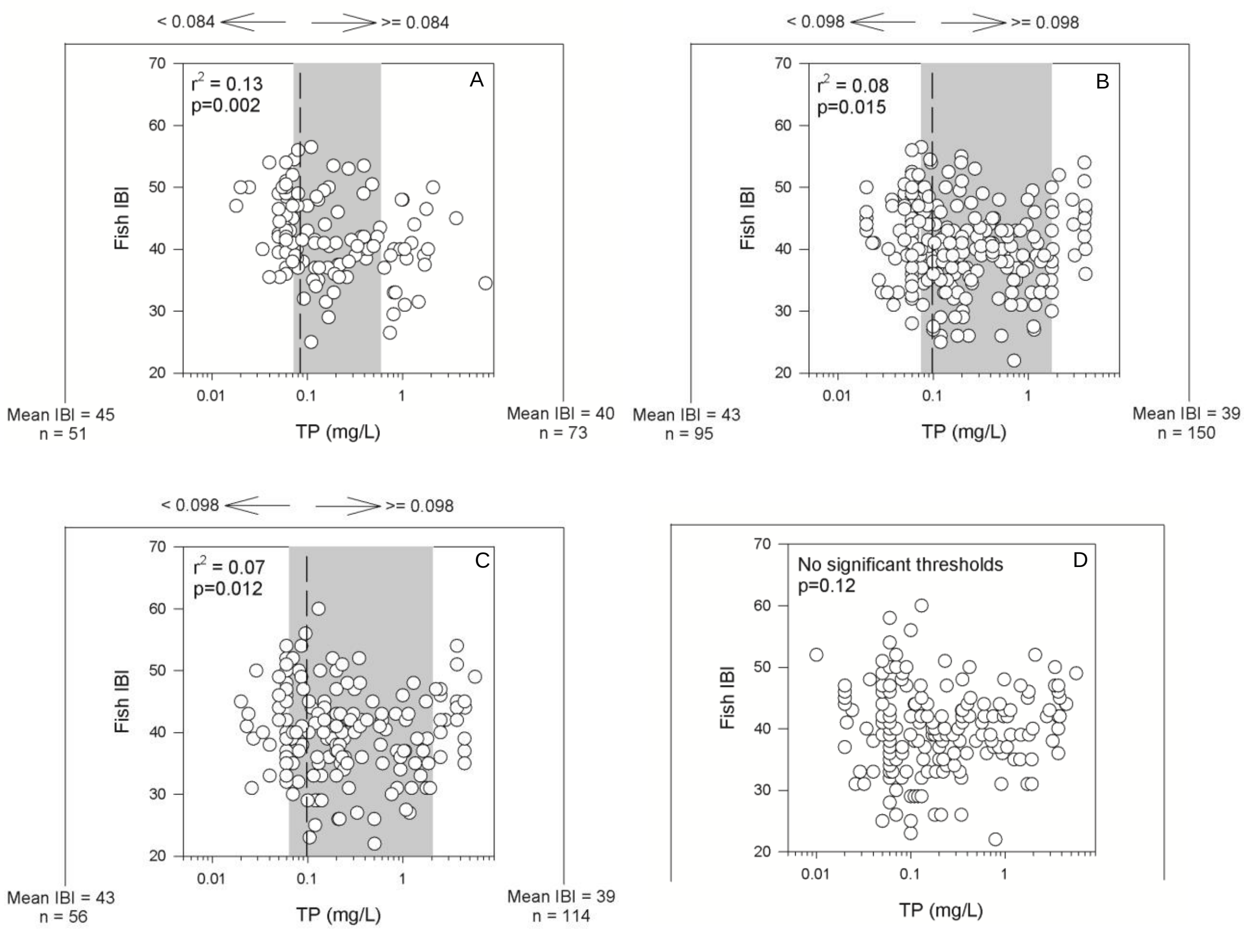


Figure 1.4.2. The relationship between Fish Index of Biotic Integrity (IBI) and total phosphorus (TP) across Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

Fish IBI also decreased as TN increased in Texas streams and rivers. The range of TN thresholds for the broadest time periods was narrow (1.2-1.4 mg/L; fig. 1.4.4.A-B), but a much higher threshold was identified for the critical period (4.8 mg/L; fig. 1.4.4C). TN was the strongest predictor for fish response for data collected over the period of record, annually, and during the index period were considered (r^2=0.13-0.22). TN models performed similarly to TP and HQI for the critical period (r^2=0.08). Thresholds ranging from 1.2-1.4 mg/L were in the upper range (0.634-1.36 mg/L) of thresholds identified by Wang et al. (2007) and Weigel and Robertson (2007) by comparing TN concentrations and Fish IBI in US streams. The critical period TN threshold of 4.8 mg/L exceeded this range by approximately 4 times, but also had a wide range of possible values within the 95% confidence range (~0.70-7.0 mg/L). No statistically significant TN threshold was identified for fish response in the paired observations dataset. As with TP, model strength and significance for comparisons of HQI and TN declined as the time frame narrowed.

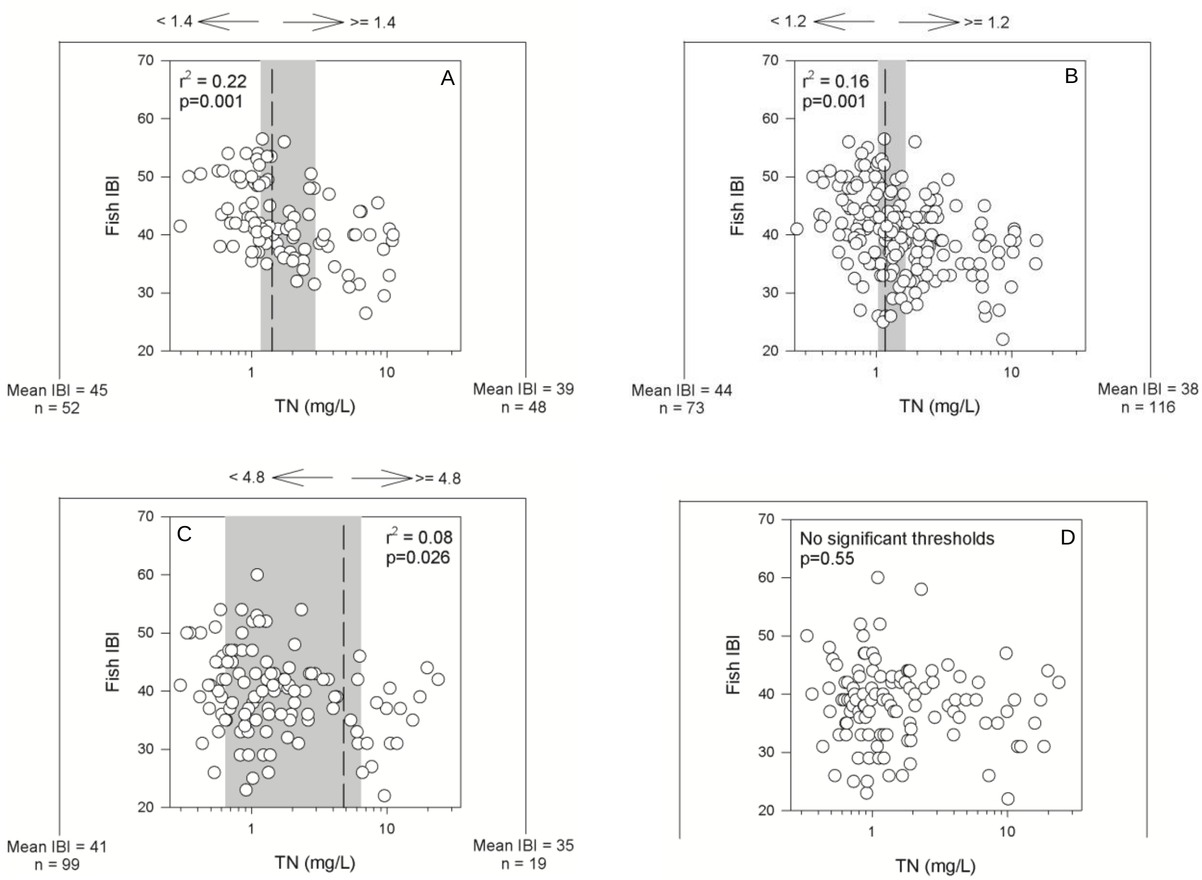


Figure 1.4.3. The relationship between Fish Index of Biotic Integrity (IBI) and total nitrogen (TN) across Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

Macrobenthic RBIBI

Both nutrients and habitat quality influence benthic macroinvertebrate populations in streams and rivers. Increased nutrient concentrations often negatively impact macroinvertebrate communities, while increased habitat quality promotes macroinvertebrate abundance and diversity. In Texas streams and rivers, RBIBI responded to nutrient and habitat variables according to established ecological theory and decreased as TP concentration increased and increased as HQI increased. HQI was consistently the strongest predictor of RBIBI for all time periods over which the bioassessment data were grouped (r^2=0.14-0.21). On average, RBIBI scores were 5-6 points greater when HQI equaled or exceeded 21, regardless of the data time frame (fig. 1.4.4).

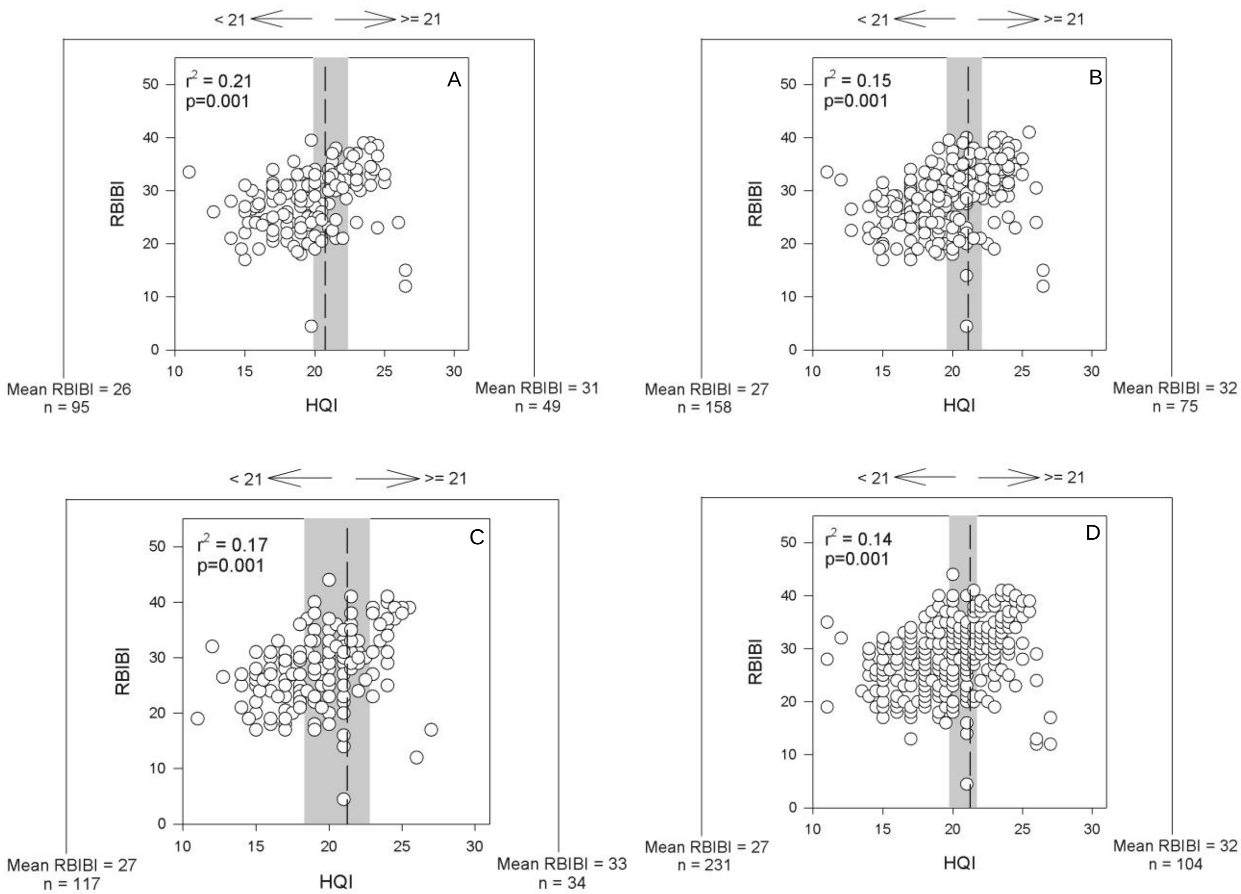


Figure 1.4.4. The relationship between the macrobenthic rapid bioassessment index of biotic integrity (RBIBI) and habitat quality index (HQI) across Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

TP thresholds varied minimally by the time period of bioassessment data considered, ranging from 0.081 mg/L annually and during the index period (r^2=0.05-0.06) to 0.11 mg/L for the period of record (r^2=0.07; fig. 1.4.5). These thresholds fall within the range of TP criteria (i.e., 0.04-0.15 mg/L) established for benthic macroinvertebrate variables in the literature (Wang et al., 2007, Weigel and Robertson, 2007, Evans-White et al., 2009). No statistically significant TP thresholds were identified for the critical period or the paired observations dataset. No statistically significant relationship between RBBI and TN concentration was found for any time period in the bioassessment database.

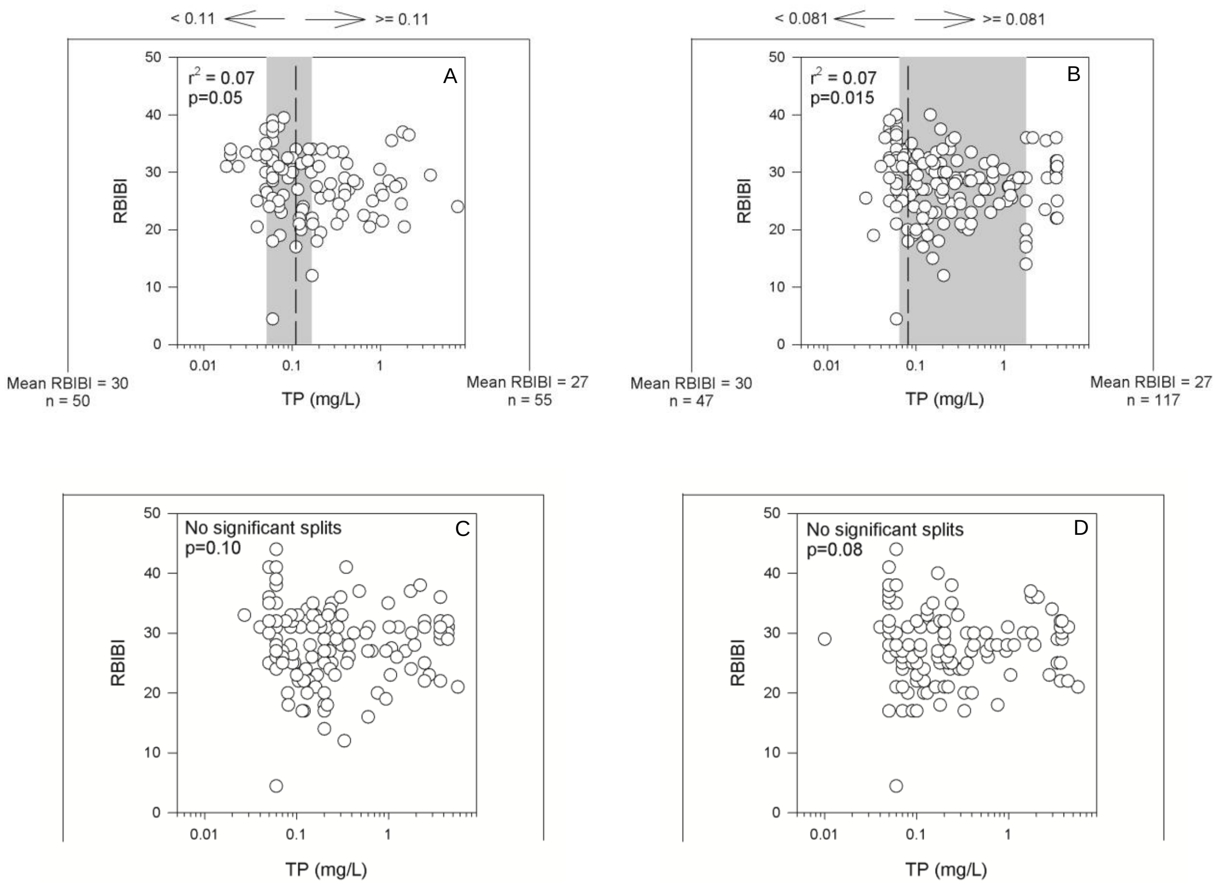


Figure 1.4.5. The relationship between the macrobenthic rapid bioassessment index of biotic integrity (RBIBI) and total phosphorus (TP) across Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

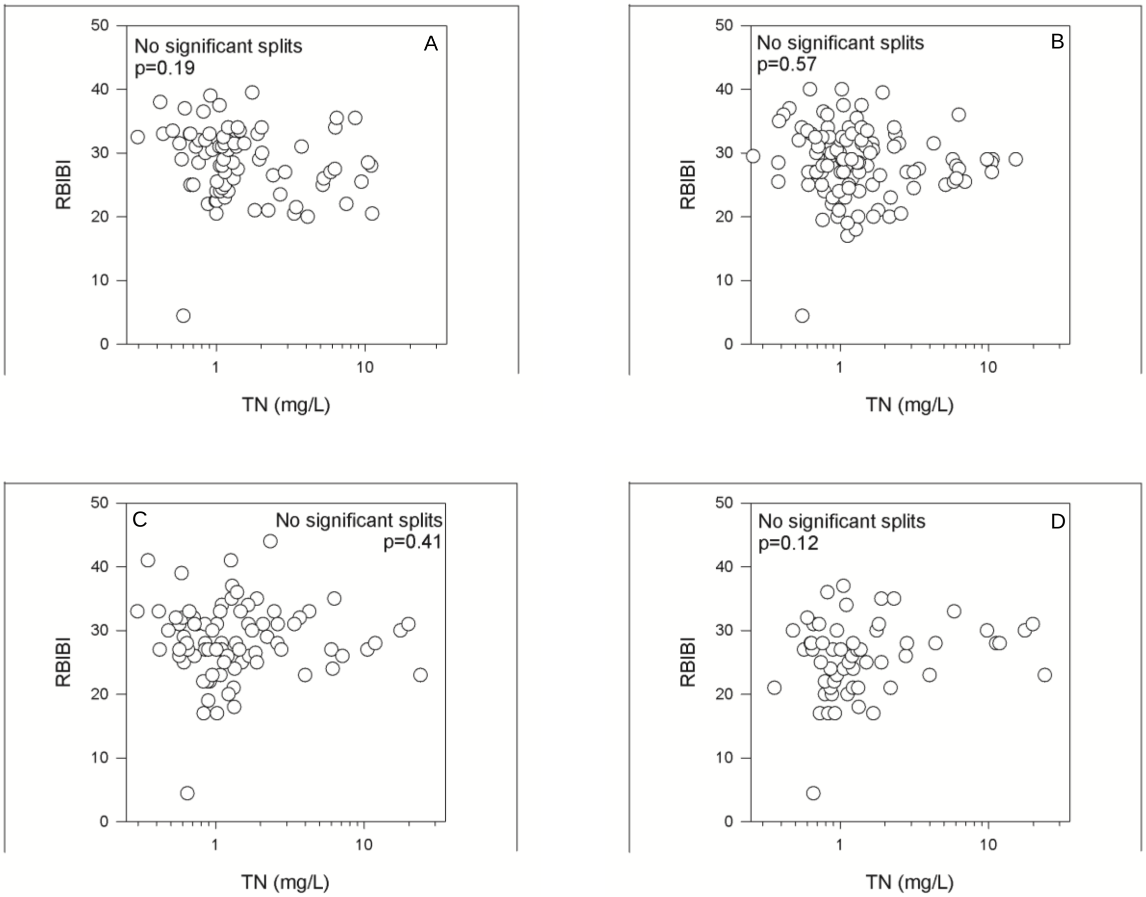


Figure 1.4.6. The relationship between the macrobenthic rapid bioassessment index of biotic integrity (RBIBI) and total nitrogen (TN) across Texas streams bioassessment datasets comprised of medians calculated for the entire period of record (A), for each year of the period of record (B), and for each critical period of the period of record (C), as well as individual paired observations (D). Thresholds were determined using classification and regression tree analysis (CART). Results for CART models on the index period dataset are included in Table 1.4.1, but were not shown graphically.

Secchi Transparency

For the stream and river bioassessment study sites, TP concentration was consistently the strongest predictor of Secchi transparency across all time periods considered (Table 1.4.1), which was consistent with findings from the larger water quality database (see Section 1.3) and with established ecological theory. Secchi transparency decreased as TP increased in Texas streams and rivers and changepoints ranged from 0.062 mg/L for the paired observations dataset to 0.089 mg/L for the entire period of record. Model strength and significance were relatively constant across time periods, but model strength was somewhat lower for the bioassessment database than for the larger water quality database (r^2=0.13-0.16 vs. 0.18-0.23). Habitat was also a predictor variable for Secchi transparency for the period of record, annual, and index period. Secchi transparency increased as HQI increased, with HQI thresholds ranging from 18 to 20. (Table 1.4.1). However, no statistically significant threshold in habitat scores was found for

Secchi transparency for the critical period or the paired observations dataset. TN was the weakest predictor of Secchi transparency; thresholds identified across the various bioassessment time periods ranged from 0.56 mg/L for the critical period to 0.71 mg/L annually. No statistically significant TN thresholds were identified for the period of record or paired observations data. Relationships between Secchi transparency and HQI, TP, and TN for the paired observations dataset are shown in Figure 1.4.7. Discussion of Secchi transparency is limited in the nutrient criteria literature, but the thresholds identified for Texas streams and rivers based on Secchi transparency values were within the range of thresholds identified for other biological response variables in this study. The TP and TN thresholds identified for the streams and rivers bioassessment database were within range of those identified for the larger water quality median database, or somewhat lower.

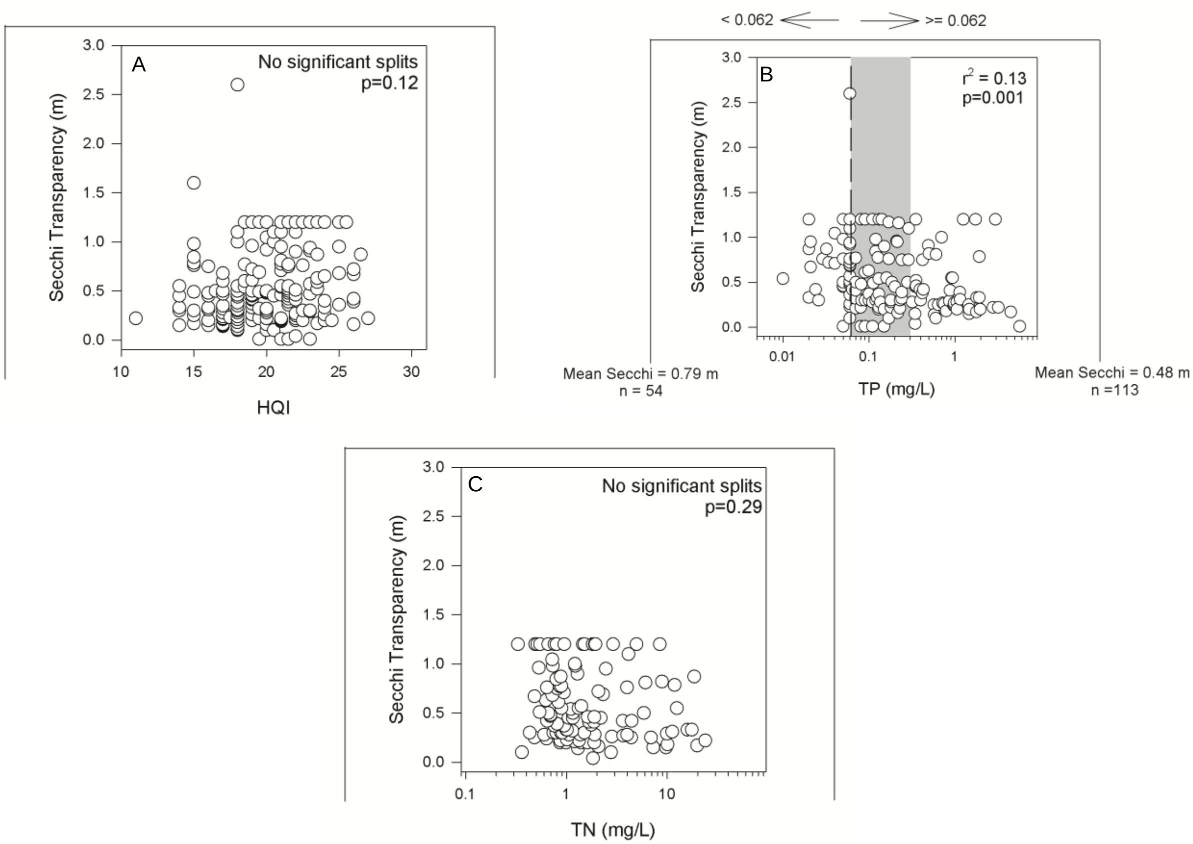


Figure 1.4.7. The relationship between Secchi transparency and habitat quality index (HQI; A), total phosphorus (TP; B), and total nitrogen (TN; C) across Texas streams bioassessment datasets comprised of medians calculated for individual paired observations showing thresholds based on classification and regression tree analysis (CART).

Sestonic Chlorophyll-a

The direct relationship between nutrients concentrations and chl-a is well established within ecological theory (Dodds et al. 2002, 2006), but was not evident in the Texas streams and rivers bioassessment database (Table 1.4.1). No statistically significant thresholds in nutrient concentrations were identified for chl-a spec or chl- a fluoro. However, for sestonic chl-a spec, statistically significant changepoints were identified in habit quality for the period of record, critical, and paired observation time frames. As HQI decreased, sestonic chl-a spec concentrations increased on average, most likely reflecting more open canopy and greater light availability for algal growth. Threshold HQI ranged from 17 for the paired observation and index period data to 20 for the period of record data, which was consistent with HQI thresholds identified for other response variables in this study. The relationship between both sestonic chl-a spec and chl-a fluoro and TP, TN, and HQI for the paired observations data are shown in Figures 1.4.8 and 1.4.9, respectively.

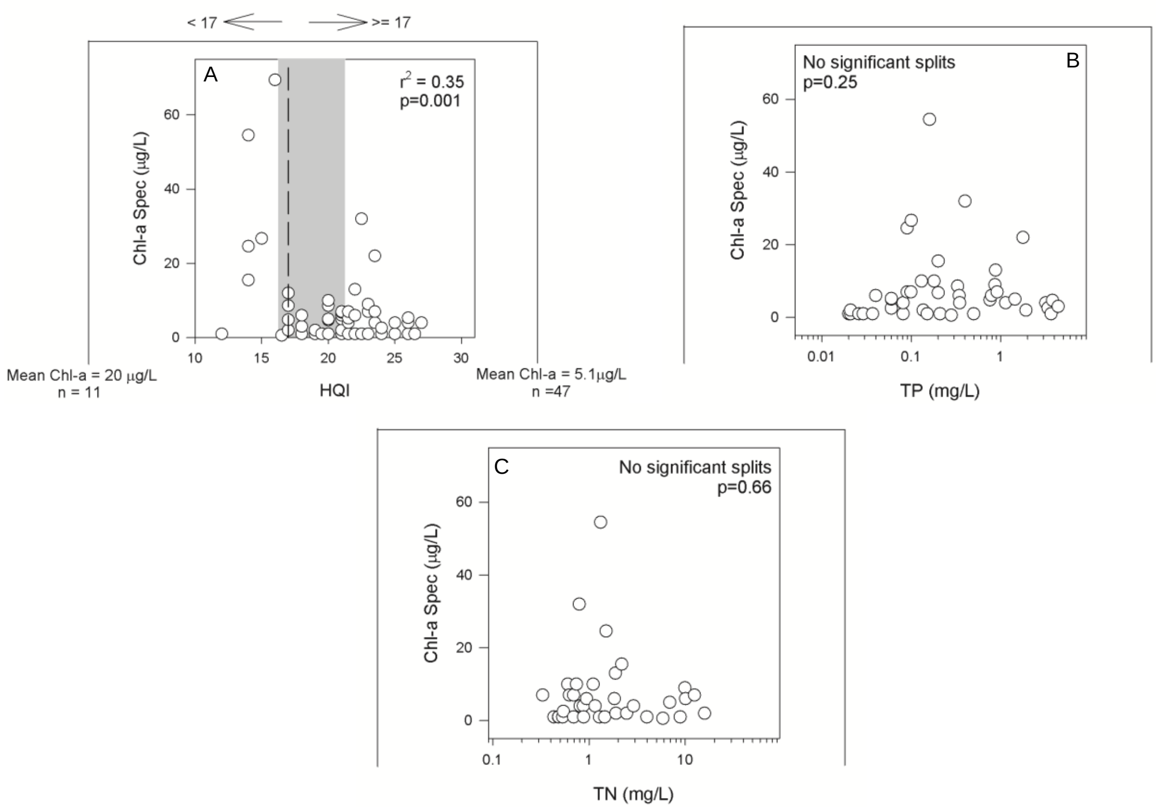


Figure 1.4.8. The relationship between sestonic spectrophotometric chlorophyll-a (chl-a spec) concentration and habitat quality index (HQI; A), total phosphorus (TP; B), and total nitrogen (TN; C) across Texas streams bioassessment datasets comprised of medians calculated for individual paired observations showing thresholds based on classification and regression tree analysis (CART).

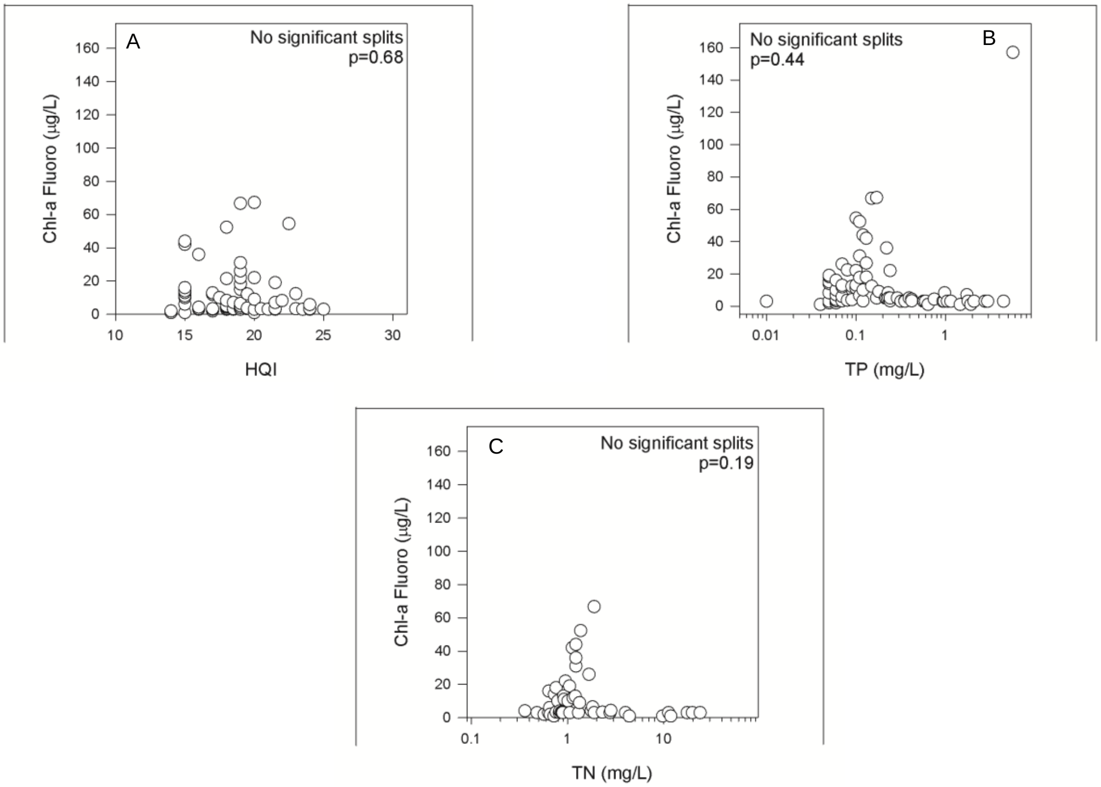


Figure 1.4.9. The relationship between sestonic fluorometric chlorophyll-a (chl-a fluoro) concentration and habitat quality index (HQI; A), total phosphorus (TP; B), and total nitrogen (TN; C) across Texas streams bioassessment datasets comprised of medians calculated for individual paired observations (A and B) and period of record (C) showing thresholds based on classification and regression tree analysis (CART).

Stressor-Response Analysis of Bioassessment Data Summary

For the fish and macrobinvertebrate indices of biotic integrity (IBIs) considered in this study, habitat quality (HQI) was consistently a strong predictor variable with threshold values ranging from 18-21. Nutrient thresholds for these variables were consistently weak or not statistically significant. These analyses could potentially be refined in the future to incorporate the categorical and/or regional nature of IBI's, which are normalized to aquatic life uses in Texas (Jill Csekitz, personal communication). For example a Fish IBI of >= 36 is considered exceptional in ecoregions 25 and 26, but intermediate in most other ecoregions. Future work could also consider individual components of IBI and HQI scores, %Canopy cover for HQI, %Tolerant Species, or %EPT taxa for RBIBI.

Table 1.4.1. Thresholds identified in classification and regression tree (CART) analyses of the streams and rivers bioassessment database. Gray text indicates a threshold that is not supported by established ecological theory. A dash indicates that no statistically significant threshold could be identified for the given stressor-response pairing. For these pairs, r^2 and p values are italicized. In the paired dataset, CART analysis did not identify models for TP or TN for subsequent testing in nCPA. Therefore, r^2 and p values were not available for these stressor-response pairings.

	Biological Variable	TP (mg/L) CP	TP r^2	TP p_{perm}	TN (mg/L) CP	TN r^2	TN p_{perm}	HQI CP	HQI r^2	HQI p_{perm}
Period of Record	Fish IBI	0.084↓	0.13	0.002	1.4↓	0.22	0.001	21↑	0.08	0.003
	RBIBI	0.11↓	0.07	0.05	--	*0.07*	*0.19*	21↑	0.21	0.001
	Secchi	0.089↓	0.16	0.002	--	*0.03*	*0.61*	20↑	0.12	0.004
	DO Flux	--	*0.03*	*0.65*	--	*0.05*	*0.42*	--	*0.07*	*0.054*
	Chl-a Spec	--	*0.06*	*0.17*	--	*0.04*	*0.54*	20↓	0.15	0.003
	Chl-a Fluoro	--	*0.03*	*0.48*	--	*0.05*	*0.31*	--	*0.03*	*0.57*
Annual	Fish IBI	0.098↓	0.08	0.001	1.2↓	0.16	0.001	18↑	0.07	0.001
	RBIBI	0.081↓	0.07	0.015	--	*0.02*	*0.57*	21↑	0.15	0.001
	Secchi	0.077↓	0.13	0.001	0.71↓	0.05	0.030	18↑	0.07	0.001
	DO Flux	0.048↓	0.06	0.044	--	*0.03*	*0.22*	--	*0.02*	*0.21*
	Chl-a Spec	--	*0.02*	*0.59*	--	*0.01*	*0.92*	--	*0.06*	*0.13*
	Chl-a Fluoro	--	*0.04*	*0.39*	--	*0.09*	*0.073*	--	*0.01*	*0.86*
Index	Fish IBI	0.096↓	0.07	0.003	1.2↓	0.13	0.002	18↑	0.07	0.001
	RBIBI	0.081↓	0.05	0.05	--	*0.04*	*0.38*	21↑	0.14	0.001
	Secchi	0.070↓	0.13	0.001	0.63↓	0.05	0.045	18↑	0.07	0.001
	DO Flux	--	*0.03*	*0.31*	--	*0.03*	*0.28*	--	*0.03*	*0.10*
	Chl-a Spec	--	*0.05*	*0.10*	--	*0.03*	*0.59*	--	*0.04*	*0.18*
	Chl-a Fluoro	--	*0.04*	*0.24*	--	*0.08*	*0.13*	--	*0.01*	*0.79*
Critical	Fish IBI	0.098↓	0.07	0.012	4.8↓	0.08	0.026	21↑	0.09	0.001
	RBIBI	--	*0.06*	*0.099*	--	*0.04*	*0.41*	21↑	0.17	0.001
	Secchi	0.069↓	0.16	0.001	0.56↓	0.08	0.027	--	*0.05*	*0.081*
	DO Flux	0.35↓	0.07	0.05	--	*0.05*	*0.23*	--	*0.01*	*0.68*
	Chl-a Spec	--	*0.07*	*0.14*	--	*0.06*	*0.33*	17↓	0.17	0.008
	Chl-a Fluoro	--	*0.02*	*0.95*	--	*0.07*	*0.31*	--	*0.03*	*0.68*
Matching	Fish IBI	--	*0.04*	*0.12*	--	*0.03*	*0.55*	18↑	0.06	0.001
	RBIBI	--	*0.06*	*0.083*	--	*0.10*	*0.12*	21↑	0.14	0.001
	Secchi	0.062↓	0.13	0.001	--	*0.04*	*0.29*	--	*0.03*	*0.12*
	DO Flux	--	--	--	--	--	--	--	*0.03*	*0.37*
	Chl-a Spec	--	*0.07*	*0.26*	--	*0.04*	*0.66*	17↓	0.35	0.001
	Chl-a Fluoro	--	*0.03*	*0.44*	--	*0.08*	*0.19*	--	*0.02*	*0.68*

↓ The value of the response variable decreases with increasing predictor variable values, and vice versa

↑ The value of the response variable increases with increasing predictor variable values, and vice versa

Model strength and significance did increase for Fish IBI vs. TN and TP and RBIBI vs. TP as the temporal scale of analysis widened from the paired observations to the period of record dataset. This finding may reflect the fact that bioassessment metrics such as Fish IBI and RBIBI were implemented as indicators of long-term nutrient concentrations and habitat conditions in streams and are potentially superior indicators of water quality to point measurements of nutrient concentrations. Poor temporal pairing of nutrient and bioassessment observations could also have played a role in this finding.

Comparisons of Secchi transparency and TP and TN concentrations yielded similar CART models in the bioassessment datasets as in the larger water quality median database, but model strength and threshold values were somewhat reduced. No statistically significant relationship between sestonic chlorophyll-a measured spectrophotometrically or fluorometrically was evident in the more limited bioassessment datasets.

REFERENCES

Archuleta, Gonzalez, and Maltby, 2012. Automated Delineation and Characterization of Watersheds for more than 3,000 Surface-Water-Quality Monitoring Stations Active in 2010 in Texas. USGS Open-File Report 2012–1077.

Ahearn, D.S., R.W. Sheibley, R.A. Dahlgren, M. Anderson, J. Johnson, and K.W. Tate. 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. J. of Hydrol. 313:234-247.

Binkley, D., G.G. Ice, J. Kaye, and C.A. Williams. 2004. Nitrogen and phosphorus concentrations in forest streams of the United States. Journal of American Water Resources Association, 40:1277-1291.Brett et al. 2005

Buck, O., D.K. Niyogi, and C.R. Townsend. 2004. Scale-dependence of land use effects on water quality of stream in agricultural catchments. Environ. Pollut. 130:287-299.

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

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

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

Dodds WK, Oakes RM. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography Methods 2:333-341.

East T L, Sharfstein B. 2006. Development of a decision tree model for the prediction of the limitation potential of phytoplankton in Lake Okeechobee, Florida, USA. Arch. Hydrobiol. 2006;165:127-144.Environmental Protection Agency (EPA). 2000. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002.

Environmental Protection Agency (EPA). 2010. Using stressor-response relationships to derive numeric nutrient criteria. EPA 820-S-10-001.

Evans-White MA, Dodds WK, Huggins DG, Baker DS. 2009. Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams. Journal of the North American Benthological Society 28:855-868.

Evans-White, M.A., B.E. Haggard, and J.T. Scott. *In Press*. A review of stream nutrient criteria development in the United States. Jounal of Environmental Quality.Gorman, O.T. and J.R. Karr. 1978. Habitat structure and stream fish communities. Ecology. 507-515.

Haggard, B.E. 2010. Phosphorus concentrations, loads and sources within the Illinois River Drainage Area, northwest Arkansas, 1997-2008. J. Environ. Qual. 39:2113-2120.

Haggard, B.E., J.T. Scott, and S.D. Longing. 2013. Sestonic chlorophyll-a shows hierarchical structure and thresholds with nutrients across the Red River Basin, USA. J. Environ. Qual. 42:437-445.

Herlihy AT, Sifneos JC. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. Journal of the North American Benthological Society 27:932-948.

Ice G, Binkley D. 2003. Forest stormwater concentrations of nitrogen and phosphorus: a comparison with EPA's proposed water quality criteria. Journal of Forestry 101:21-28.

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(3):998-1008.

King RS, and CJ Richardson. 2003. Integrating bioasessment and ecological risk assessment: An approach to developing numerical water-quality critera. Environmental Management. 31:795-809. doi:10.1007/s00267-002-0036-4Longing SD, Haggard BE. 2010. Distributions of median nutrient and chlorophyll concentrations across the Red River Basin, USA. Journal of Environmental Quality 39:1966-1974.

Qian SS, King RS, Richardson CJ. 2003. Two methods for the detection of environmental thresholds. Ecological Modeling 166:87-97.Robertson et al. 2006

Rohm CM, Omernik JM, Woods AJ, Stoddard JL. 2002. Regional characteristics of nutrient concentrations in streams and their application to nutrient criteria development. Journal of the American Water Resources Association 38:213-239.

Royer TV, David MB, Gentry LE, Mitchell CA, Starks KM, Heatherly T, Whiles MR. 2008. Assessment of chlorophyll-a as a criterion for establishing nutrient standards in streams and rivers of Illinois. Journal of Environmental Quality 37:437447.

Smith RA, Alexander RB, Schwarz GE. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous Unites States. Environmental Science and Technology 37:3039-3047.

Stevenson RJ, Hill BH, Herlihy AT, Yuan LL, Norton SB. 2008. Algae-P relationships, thresholds, and frequency distributions guide nutrient criterion development. Journal of the North American Benthological Society 27:783-799

Suplee MW, Varghese A, Cleland J. 2007. Developing nutrient criteria for streams: an evaluation of the frequency distribution method. Journal of the American Water Resources Association 43(2):453-472.

Urban, D.L. (2002) "Classification and regression trees" In: McCune, B., Grace, J.B. eds. , Analysis of Ecological Communities, MjM Software Design, Gleneden Beach, Oregon, USA, pp 221-231.

Wang LZ, Robertson DM, Garrison PJ. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. Environmental Management 39:194-212.

Weigel BM, Robertson DM. 2007. Identifying biotic integrity and water chemistry relationships in nonwadeable rivers of Wisconsin: toward the development of nutrient criteria. Environmental Management 40:691-708.

APPENDICES

Appendix 1.1. This Macro is used before the pivot table is created; the Macro inserts dummy data into the basin data worksheet so that after each station is pivoted, a column exists for each TCEQ parameter code (n=116).

```
Code:  Sub Insert()
    Dim r As Long
    Dim mcol As String
    Dim i As Long
    Application.ScreenUpdating = True
    Application.Calculation = xlCalculationManual
     ' find last used cell in Column E
    r = Cells(Rows.Count, "E").End(xlUp).Row
     ' get value of  last used cell in column E
    mcol = Cells(r, 5).Value
     ' insert rows by looping from bottom
      For i = r To 2 Step -1
        If Cells(i, 5).Value <> mcol Then
            Sheet2.Range("E2:E119").Value = mcol
            Sheet2.Range("B2:B119").Value = Cells(i + 1, 2).Value
            Sheet2.Range("A2:A119").Value = Cells(i, 1).Value
            mcol = Cells(i, 5).Value
            Sheet2.Range("A2:U119").Copy
            Sheet1.Cells(i + 1, 1).Insert Shift:=xlDown
           ' Clears any added content in Dummy Data Tab
           Sheet2.Range("A2:E119").ClearContents
        End If
    Next i
    Application.Calculation = xlCalculationAutomatic
End Sub
```

Appendix 1.2. This Macro is used after the pivot table has been created for each Station ID; the Macro changes the font color of data that is flagged with a > or < sign to red and deletes the column containing the flag (e.g., >).

```
Code: Sub Edit()

   Dim counter As Integer
   counter = 0
   Dim x As Integer

   Sheets("Copy Here").Select
   Range("A3").Select

   Do While ActiveCell.Value <> ""
      ActiveCell.Offset(1, 0).Select
      counter = counter + 1
   Loop

   columnnum = 2
   Range("B3").Select

   Do While ActiveCell.Value <> ""
      If ActiveCell.Offset(1, 0).Value <> 1 Then
         Columns(columnnum).Select
         Selection.Delete Shift:=xlToLeft
         Cells(3, columnnum).Select
      End If
      For x = 1 To counter
         If IsError(ActiveCell.Offset(1, 1).Value) Then
            ActiveCell.Offset(1, 0).Select
               With Selection.Font
                  .Color = -16776961
                  .TintAndShade = 0
               End With
         Else
            ActiveCell.Offset(1, 0).Select
          End If
      Next x
      Cells(3, columnnum).Select
      ActiveCell.Offset(0, 1).Select
      columnnum = columnnum + 1
   Loop

   Cells.Select
   Selection.Copy
```

```
    Sheets("Results").Select
    Range("A1").Select
    ActiveSheet.Paste

End Sub
```

Appendix 1.3. This Macro is used after the pivot table has been created for each Station ID and data has been flagged; the Macro inserts new columns that contain the calculated parameters.

```
Code:  Sub Edit()
    Dim i As Integer
    For i = 1 To ThisWorkbook.Sheets.Count
      Set ws = Worksheets(i)
       Worksheets(i).Activate


        If Not ws.Name = "Pivot" _
        And Not ws.Name = "Basin" _
        And Not ws.Name = "Parameter Code Description" Then

         'Deletes Row if Column Labels is in the first row
         If Range("B1").Value = "Column Labels" Then
           Rows(1).Delete
         End If


        Columns(4).Insert Shift:=xlToRight
        Columns(16).Insert Shift:=xlToRight
        Columns(18).Insert Shift:=xlToRight
        Columns(19).Insert Shift:=xlToRight
        Columns(26).Insert Shift:=xlToRight
        Columns(30).Insert Shift:=xlToRight
        Columns(33).Insert Shift:=xlToRight


          ' Inserts Columns for 00600 group
            For colx = 59 To 66 Step 1
              Columns(colx).Insert Shift:=xlToRight
             Next

         Columns(72).Insert Shift:=xlToRight
         Columns(76).Insert Shift:=xlToRight
         Columns(77).Insert Shift:=xlToRight
         Columns(78).Insert Shift:=xlToRight
         Columns(84).Insert Shift:=xlToRight
         Columns(99).Insert Shift:=xlToRight
         Columns(107).Insert Shift:=xlToRight
         Columns(139).Insert Shift:=xlToRight
         Columns(142).Insert Shift:=xlToRight

         'Names inserted Columns
```

```
Cells(1, 4).Value = "00010C"
Cells(1, 16).Value = "00077m"
Cells(1, 18).Value = "00078C1"
Cells(1, 19).Value = "00078C"
Cells(1, 26).Value = "00210C"
Cells(1, 30).Value = "00213C"
Cells(1, 33).Value = "00215C"
Cells(1, 59).Value = "00593C1"
Cells(1, 60).Value = "00593C2"
Cells(1, 61).Value = "00600A"
Cells(1, 62).Value = "00600B"
Cells(1, 63).Value = "00600i"
Cells(1, 64).Value = "00600C1"
Cells(1, 65).Value = "00600C2"
Cells(1, 66).Value = "00600C"
Cells(1, 72).Value = "00620C1"
Cells(1, 76).Value = "00630C"
Cells(1, 77).Value = "00630C1"
Cells(1, 78).Value = "00630C2"
Cells(1, 84).Value = "00671C"
Cells(1, 99).Value = "20389C"
Cells(1, 107).Value = "20485C"
Cells(1, 139).Value = "89077m"
Cells(1, 142).Value = "89856C"


    Dim LastRow As Long
    LastRow = Cells(Rows.Count, "A").End(xlUp).Row
      If LastRow < 4 Then On Error Resume Next

     'Runs Equation for Column 00010C
      Range("D3").Select
      ActiveCell.FormulaR1C1 = "=IF(RC[1]="""",RC[-1],((RC[1]-32)*(5/9)))"
      Range("D3").Select
      Selection.AutoFill Destination:=Range("D3:D" & LastRow)

     'Runs Equation for Column 00077m
      Range("P3").Select
      ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""","""",(RC[-1]/39.700787))"
      Range("P3").Select
      Selection.AutoFill Destination:=Range("P3:P" & LastRow)

     'Runs Equation for Column 00078C1
      Range("R3").Select
      ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[-2],RC[-1])"
```

```
    Range("R3").Select
    Selection.AutoFill Destination:=Range("R3:R" & LastRow)

'Runs Equation for Column 00078C
    Range("S3").Select
    ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[120],RC[-1])"
    Range("S3").Select
    Selection.AutoFill Destination:=Range("S3:S" & LastRow)

'Runs Equation for Column 00210C
    Range("Z3").Select
    ActiveCell.FormulaR1C1 = "=IF(AND(RC[-1]>0,RC[1]>0),RC[-1]-RC[1],"""")"
    Range("Z3").Select
    Selection.AutoFill Destination:=Range("Z3:Z" & LastRow)

'Runs Equation for Column 00213C
    Range("AD3").Select
    ActiveCell.FormulaR1C1 = "=IF(AND(RC[-1]>0,RC[1]>0),RC[-1]-RC[1],"""")"
    Range("AD3").Select
    Selection.AutoFill Destination:=Range("AD3:AD" & LastRow)

'Runs Equation for Column 00215C
    Range("AG3").Select
    ActiveCell.FormulaR1C1 = "=IF(AND(RC[-1]>0,RC[1]>0),RC[-1]-RC[1],"""")"
    Range("AG3").Select
    Selection.AutoFill Destination:=Range("AG3:AG" & LastRow)

'Runs Equation for Column 00593C1
    Range("BG3").Select
    ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[13],RC[-1])"
    Range("BG3").Select
    Selection.AutoFill Destination:=Range("BG3:BG" & LastRow)

'Runs Equation for Column 00593C2
    Range("BH3").Select
    ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[12],RC[-1])"
    Range("BH3").Select
    Selection.AutoFill Destination:=Range("BH3:BH" & LastRow)

'Runs Equation for Column 00600A
    Range("BI3").Select
    ActiveCell.FormulaR1C1 = "=IF(AND(RC[13]>0,RC[14]>0),RC[13]+RC[14],"""")"
    Range("BI3").Select
    Selection.AutoFill Destination:=Range("BI3:BI" & LastRow)
```

```
'Runs Equation for Column 00600B
  Range("BJ3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[12]>0,RC[-4]>0),RC[12]+RC[-4],"""")"
  Range("BJ3").Select
  Selection.AutoFill Destination:=Range("BJ3:BJ" & LastRow)

'Runs Equation for Column 00600i
  Range("BK3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[11]>0,RC[8]>0,RC[7]>0),RC[11]+RC[8]+RC[7],"""")"
  Range("BK3").Select
  Selection.AutoFill Destination:=Range("BK3:BK" & LastRow)

'Runs Equation for Column 00600C1
  Range("BL3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-3]="""",RC[-2],RC[-3])"
  Range("BL3").Select
  Selection.AutoFill Destination:=Range("BL3:BL" & LastRow)

'Runs Equation for Column 00600C2
  Range("BM3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-3]="""",RC[-2],RC[-3])"
  Range("BM3").Select
  Selection.AutoFill Destination:=Range("BM3:BM" & LastRow)

'Runs Equation for Column 00600C
  Range("BN3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-2]="""",RC[-1],RC[-2])"
  Range("BN3").Select
  Selection.AutoFill Destination:=Range("BN3:BN" & LastRow)

'Runs Equation for Column 00620C1
  Range("BT3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[-2]>0,RC[-1]>0),RC[-2]+RC[-1],"""")"
  Range("BT3").Select
  Selection.AutoFill Destination:=Range("BT3:BT" & LastRow)

'Runs Equation for Column 00630C
  Range("BX3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[2]>0,RC[2],RC[-16])"
  Range("BX3").Select
  Selection.AutoFill Destination:=Range("BX3:BX" & LastRow)

'Runs Equation for Column 00630C1
  Range("BY3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-2]="""",RC[-19],RC[-2])"
```

```
  Range("BY3").Select
  Selection.AutoFill Destination:=Range("BY3:BY" & LastRow)

'Runs Equation for Column 00630C2
  Range("BZ3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[-19],RC[-1])"
  Range("BZ3").Select
  Selection.AutoFill Destination:=Range("BZ3:BZ" & LastRow)

'Runs Equation for Column 00671C
  Range("CF3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""",RC[43],RC[-1])"
  Range("CF3").Select
  Selection.AutoFill Destination:=Range("CF3:CF" & LastRow)

'Runs Equation for Column 20389C
  Range("CU3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[-1]>0,RC[-2]>0),RC[-1]-RC[-2],"""")"
  Range("CU3").Select
  Selection.AutoFill Destination:=Range("CU3:CU" & LastRow)

'Runs Equation for Column 20485C
  Range("DC3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[1]>0,RC[-1]>0),RC[1]-RC[-1],"""")"
  Range("DC3").Select
  Selection.AutoFill Destination:=Range("DC3:DC" & LastRow)

'Runs Equation for Column 89077m
  Range("EI3").Select
  ActiveCell.FormulaR1C1 = "=IF(RC[-1]="""","""",(RC[-1]/3.2808399))"
  Range("EI3").Select
  Selection.AutoFill Destination:=Range("EI3:EI" & LastRow)

'Runs Equation for Column 89856C
  Range("EL3").Select
  ActiveCell.FormulaR1C1 = "=IF(AND(RC[-1]>0,RC[-2]>0),RC[-1]-RC[-2],"""")"
  Range("EL3").Select
  Selection.AutoFill Destination:=Range("EL3:EL" & LastRow)

  'Replaces all zeros with blank cell
    Dim rng As Range
     For Each rng In Range("B3:EN" & LastRow)
       If rng.Value = 0 Then
         rng.Value = ""
       End If
```

```
                    Next

                'Bolds completed parameters
                 Range("D1,S1,Z1,AD1,AG1,BI1,BJ1,BK1,BN1,BX1,CF1,CU1,DC1,EL1").Font.Bold = True

            Range("A1").EntireColumn.AutoFit
            Range("A1").Select
    End If
  Next

  'Returns to Pivot tab after all sheets have been edited
  Sheets("Pivot").Select
End Sub
```

Appendix 1.4. This Macro was applied to each Station workbook and was used to insert parameter code descriptions, calculate the number of data points for a given parameter and the median value of the data points.

```
Code:  Sub Median_Calculation()
Dim i As Integer
    For i = 1 To ThisWorkbook.Sheets.Count
      Set ws = Worksheets(i)
       Worksheets(i).Activate
    'These worksheets are excluded from Macro Calculations, _
        Must be sure that Sheet names are spelled exactly as _
        they appear here
      If Not ws.Name = "Pivot" _
      And Not ws.Name = "Basin 1" _
      And Not ws.Name = "Parameter Code Description" Then

              Range("A2:EN2").ClearContents
              Rows("1:2").Insert
              Rows("4:5").Insert

             'Counts the number of Columns and Rows in each sheet
             Dim LastRow As Long
              Dim LastCol As Long
              LastRow = Cells(Rows.Count, "A").End(xlUp).Row
              LastCol = Cells(3, Columns.Count).End(xlToLeft).Column

            'Copy and pastes the parameter descriptions in _
                each station sheet
            Sheets("Parameter Code Description").Select
            Range("A1:EN2").Select
            Selection.Copy
            Worksheets(i).Select
            Range("A1").Select
            ActiveSheet.Paste

            'Row names for each parameter code
            Range("A3").Value = "Code"
            Range("A4").Value = "Count,Cells"
            Range("A5").Value = "Count,Values"
            Range("A6").Value = "Median"

            'Formula for Count,Cells
            Range("B4").Select
            ActiveCell.Formula = "=COUNTA(R[3]C:R[" & LastRow & "]C)"
            Range("B4").Select
```

```
        Selection.AutoFill Destination:=Range(Cells(4, 2), Cells(4, LastCol))

        'Formula for Count, Values
        Range("B5").Select
        ActiveCell.Formula = "=COUNT(R[2]C:R[" & LastRow & "]C)"
        Range("B5").Select
        Selection.AutoFill Destination:=Range(Cells(5, 2), Cells(5, LastCol))

        'Formula for Median
        Range("B6").Select
        ActiveCell.Formula = "=IF(R[-1]C>0,MEDIAN(R[1]C:R[" & LastRow & "]C), """")"
        Range("B6").Select
        Selection.AutoFill Destination:=Range(Cells(6, 2), Cells(6, LastCol))

            'Fills Count and Median cell values blue
           Range("A4:" & ActiveSheet.Cells(6, LastCol).Address).Select
            With Selection.Interior
             .Pattern = xlSolid
             .PatternColorIndex = xlAutomatic
             .ThemeColor = xlThemeColorAccent5
             .TintAndShade = 0.799981688894314
             .PatternTintAndShade = 0

           End With

     Range("A1").Select

   End If
 Next
       'Returns to Pivot tab after Median has been calculated in all sheets
   Sheets("Pivot").Select
End Sub
```

Appendix 1.5. This Macro is used to combine the medians calculated in each individual station worksheet into the final database that contains all medians for all Station IDs. Because we wanted to be able to adjust the number of data points required to calculate a median value, the final database was achieved using three macros. The first macro copies the count of median from each Station worksheet and inserts the counts into the final median database worksheet.

```
Code: Sub Median_Database_Count()

'This Macro takes the median from each station tab in the current Basin workbook
'And inserts it into the Median Database workbook

 Dim count As Integer
 Dim CountWs As Worksheet
 Dim BasinWs As Worksheet


 'Counts how many sheets are in the Basin workbook
 count = ThisWorkbook.Sheets.count

 'this defines the Sheet "Count" in the Median Database
 Set CountWs = Workbooks("Median Database").Sheets("Count")

 'this defines the Sheet "Basin" in the Basin workbook
 Set BasinWs = ThisWorkbook.Sheets("Basin")

 'Inserts blank rows in Median Database
 CountWs.Activate
 Range("B4:B" & count).Select
 Selection.EntireRow.Insert

 Dim i As Integer
  For i = 1 To ThisWorkbook.Sheets.count
   Set ws = ThisWorkbook.Worksheets(i)
    ws.Activate

        'Identifies sitenum with each station tab name in the Basin workbook
         Dim sitenum As String
          sitenum = ws.Name

      If Not ws.Name = "Pivot" _
      And Not ws.Name = "Basin" _
      And Not ws.Name = "Parameter Code Description" Then

        'places the station tab name in the Median Database
```

```
        Dim x As Integer
          For Each ws In Worksheets
            For x = 4 To i
              If CountWs.Cells(x, 2).Value = "" Then
                CountWs.Cells(x, 2).Value = sitenum
              End If
             'Places "Cell,Value" row in Median Database
              If CountWs.Cells(x, 3).Value = "" Then
                Range("B5:EM5").Select
                Selection.Copy
                CountWs.Range("C" & x).PasteSpecial xlPasteValues
                Range("A1").Select
               End If
            Next x
          Next ws

       'Places the Basin number in the Median Database
        Dim r As Integer
        BasinNum = BasinWs.Cells(2, 1).Value
          For r = 4 To i
            If CountWs.Cells(r, 2).Value <> "" Then
              CountWs.Cells(r, 1).Value = BasinNum
            End If
          Next r

  'ends the If Not command
  End If
'moves to next "i"
Next

 'Returns Basin workbook back to Pivot sheet
 Sheets("Pivot").Select
 'Selects cell "A3" in Median Database
 CountWs.Activate

 Range("C4:EN" & count).Font.Bold = False

'Sorts data in Median Database by Basin number
 Dim LastRow As Long
 LastRow = Cells(Rows.count, "A").End(xlUp).Row
   Range("A4:EN" & LastRow).Select
   CountWs.Sort.SortFields.Clear
   CountWs.Sort.SortFields.Add Key:=Range("A4:A" & LastRow) _
      , SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:=xlSortNormal
    With CountWs.Sort
```

```
        .SetRange Range("A3:EN" & LastRow)
        .Header = xlYes
        .MatchCase = False
        .Orientation = xlTopToBottom
        .SortMethod = xlPinYin
        .Apply
      End With

    'Selects cell "A3" in Median Database
    Range("A3").Select

End Sub
```

The second macro takes the median from each station worksheet and inserts it into the final median database workbook.

```
Code: Sub Median_Database_Value()
'This Macro takes the median from each station tab in the current Basin workbook
'And inserts it into the Median Database workbook
  Dim count As Integer
  Dim ValueWs As Worksheet
  Dim BasinWs As Worksheet
  'Counts how many sheets are in the Basin workbook
  count = ThisWorkbook.Sheets.count
  'this defines the Sheet "Count" in the Median Database
  Set ValueWs = Workbooks("Median Database").Sheets("Value")

  'this defines the Sheet "Basin" in the Basin workbook
  Set BasinWs = ThisWorkbook.Sheets("Basin")

  'Inserts blank rows in Median Database
  ValueWs.Activate
  Range("B4:B" & count).Select
  Selection.EntireRow.Insert

  Dim i As Integer
   For i = 1 To ThisWorkbook.Sheets.count
    Set ws = ThisWorkbook.Worksheets(i)
      ws.Activate

        'Identifies sitenum with each station tab name in the Basin workbook
         Dim sitenum As String
          sitenum = ws.Name

      If Not ws.Name = "Pivot" _
```

```
    And Not ws.Name = "Basin" _
    And Not ws.Name = "Parameter Code Description" Then

      'places the station tab name in the Median Database
       Dim x As Integer
         For Each ws In Worksheets
           For x = 4 To i
             If ValueWs.Cells(x, 2).Value = "" Then
               ValueWs.Cells(x, 2).Value = sitenum
             End If
            'Places "Cell,Value" row in Median Database
             If ValueWs.Cells(x, 3).Value = "" Then
               Range("B6:EM6").Select
               Selection.Copy
               ValueWs.Range("C" & x).PasteSpecial xlPasteValues
               Range("A1").Select
              End If
              If ValueWs.Cells(x, 3).Value = "" Then
                ValueWs.Cells(x, 3).Value = 0
              End If
           Next x
         Next ws

      'Places the Basin number in the Median Database
       Dim r As Integer
       BasinNum = BasinWs.Cells(2, 1).Value
         For r = 4 To i
           If ValueWs.Cells(r, 2).Value <> "" Then
             ValueWs.Cells(r, 1).Value = BasinNum
           End If
         Next r

   'ends the If Not command
   End If
  'moves to next "i"
  Next

  'Returns Basin workbook back to Pivot sheet
  Sheets("Pivot").Select

  'Turns all zeros to blank cell in column C
  ValueWs.Activate
  Dim rng As Range
   For Each rng In Range("C4:C" & count)
     If rng.Value = 0 Then
```

```
        rng.Value = ""
      End If
    Next rng

    Range("C4:EN" & count).Font.Bold = False

   'Sorts data in Median Database by Basin number
    Dim LastRow As Long
    LastRow = Cells(Rows.count, "A").End(xlUp).Row
      Range("A4:EN" & LastRow).Select
      ValueWs.Sort.SortFields.Clear
      ValueWs.Sort.SortFields.Add Key:=Range("A4:A" & LastRow) _
         , SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:=xlSortNormal
       With ValueWs.Sort
         .SetRange Range("A3:EN" & LastRow)
         .Header = xlYes
         .MatchCase = False
         .Orientation = xlTopToBottom
         .SortMethod = xlPinYin
         .Apply
       End With

    'Selects cell "A3" in Median Database
    Range("A3").Select
End Sub
```

The final macro displays the median value for each parameter and Site ID based on the number of data point required to calculate a median (n is user set in the spreadsheet).

```
Code: Sub Median_Database_Median()
'This macro displays the median for each site where the sample number
'is greater than the value in cel B1
Dim LastRow As Long
Dim LastValue As Long
'Counts the number of cells used in the Value sheet
LastValue = Sheets("Value").Cells(Rows.Count, "A").End(xlUp).Row

'Copies the Basin # and the Site # from the Value sheet and pastes
'into the Median sheet
Sheets("Value").Select
Range("A4:B" & LastValue).Select
Selection.Copy
Sheets("Median").Range("A4").PasteSpecial xlPasteValues
Sheets("Value").Range("A3").Select
```

```
Sheets("Median").Activate

LastRow = Sheets("Median").Cells(Rows.Count, "A").End(xlUp).Row

'Equation that displays the median based on B1 value
Range("C4").Select
ActiveCell.FormulaR1C1 = "=IF(Count!RC>R1C2,Value!RC,"""")"
Range("C4").Select
Selection.Copy Destination:=Range("C4:EN" & LastRow)

End Sub
```

Appendix 1.6. General stream condition frequency distributions for Basin x level III and level IV Ecoregions.

Table A1.6.1: Frequency Distribution of Median Nutrient and Chlorophyll-a Concentrations among Basin by Level III Ecoregions in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26-Southwestern Tablelands	16	0.060	0.063	0.078	0.088	0.130	0.250	0.508
2-26-Southwestern Tablelands	18	0.020	0.040	0.043	0.060	0.060	0.375	1.010
2-27-Central Great Plains	24	0.050	0.060	0.074	0.135	0.245	0.502	1.300
2-29-Cross Timbers	2	0.200	--	--	0.205	--	--	0.210
2-32-Texas Blackland Prairies	8	0.060	0.074	0.133	0.238	2.608	3.710	4.200
2-33-East Central Texas Plains	4	0.120	--	0.140	0.169	0.381	--	0.950
2-35-South Central Plains	7	0.100	0.106	0.113	0.120	0.135	0.223	0.333
3-32-Texas Blackland Prairies	4	0.060	--	0.060	0.083	0.293	--	0.855
3-33-East Central Texas Plains	6	0.095	--	0.266	0.445	0.834	--	1.680
3-35-South Central Plains	5	0.115	--	0.140	0.190	0.210	--	0.392
4-33-East Central Texas Plains	0	--	--	--	--	--	--	--
4-35-South Central Plains	31	0.023	0.050	0.080	0.100	0.140	0.190	1.290
5-32-Texas Blackland Prairies	2	0.190	--	--	0.218	--	--	0.246
5-33-East Central Texas Plains	1	--	--	--	0.160	--	--	--
5-34-Western Gulf Coastal Plain	2	0.060	--	--	0.068	--	--	0.075
5-35-South Central Plains	25	0.060	0.060	0.060	0.100	0.120	0.146	0.210
6-33-East Central Texas Plains	0	--	--	--	--	--	--	--
6-35-South Central Plains	72	0.060	0.060	0.060	0.098	0.170	0.259	3.300
7-34-Western Gulf Coastal Plain	6	0.110	--	0.162	0.175	0.218	--	0.345
8-29-Cross Timbers	1	--	--	--	0.285	--	--	--
8-30 Edwards Plateau	75	0.050	0.060	0.060	0.065	0.133	0.910	2.880
8-32-Texas Blackland Prairies	7	0.060	0.417	0.733	0.920	1.020	1.114	1.165
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--
8-34 West Gulf Plain	6	0.060	--	0.129	0.140	0.140	--	0.960
8-35-South Central Plains	19	0.029	0.060	0.065	0.080	0.210	0.896	1.880
9-34-Western Gulf Coastal Plain	2	0.125	--	--	0.138	--	-	0.150
10-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
10-34- Western Gulf Coastal Plain	116	0.060	0.130	0.338	0.970	1.386	1.920	3.280
10-35-South Central Plains	36	0.040	0.063	0.084	0.150	0.360	1.508	3.285
11-34-Western Gulf Coastal Plain	37	0.080	0.096	0.150	0.270	0.620	0.770	0.980
12-25-High Plains	2	0.145	--	--	0.188	--	--	0.230
12-26-Southwestern Tablelands	3	0.060	--	--	0.080	--	--	1.130
12-27-Central Great Plains	10	0.060	0.078	0.091	0.400	1.055	2.080	2.975
12-29-Cross Timbers	81	0.040	0.060	0.060	0.082	0.220	0.617	1.960
12-30-Edwards Plateau	7	0.050	0.050	0.055	0.060	0.060	0.060	0.060
12-32-Texas Blackland Prairies	24	0.050	0.050	0.058	0.128	0.286	0.760	1.525
12-33-East Central Texas Plains	18	0.055	0.074	0.083	0.141	1.453	2.946	7.430
12-34-Western Gulf Coastal Plain	7	0.130	0.181	0.233	0.272	0.535	1.152	1.710
13-34-Western Gulf Coastal Plain	11	0.070	0.131	0.179	0.196	0.223	0.240	0.390
14-25-High Plains	1	--	--	--	2.230	--	--	--
14-26-Southwestern Tablelands	7	0.060	0.060	0.065	0.120	0.211	0.619	1.125
14-27-Central Great Plains	22	0.060	0.060	0.060	0.060	0.069	0.088	0.131
14-29-Cross Timbers	7	0.060	0.060	0.060	0.060	0.126	0.615	1.300
14-30-Edwards Plateau	52	0.020	0.020	0.050	0.060	0.060	0.060	1.895
14-32-Texas Blackland Prairies	19	0.020	0.020	0.035	0.060	0.153	0.218	0.484

14-33-East Central Texas Plains	8	0.060	0.060	0.060	0.258	0.370	0.376	0.390
14-34-Western Gulf Coastal Plain	5	0.125	--	0.274	0.300	0.328	--	0.364
15-34-Western Gulf Coastal Plain	1	--	--	--	0.370	--	--	--
16-32-Texas Blackland Prairies	2	0.085	--	--	0.153	--	--	0.220
16-33-East Central Texas Plains	2	0.200	--	--	0.215	--	--	0.230
16-34-Western Gulf Coastal Plain	6	0.100	--	0.174	0.198	0.278	--	0.310
18-30-Edwards Plateau	24	0.007	0.011	0.019	0.050	0.050	0.050	0.060
18-32-Texas Blackland Prairies	29	0.050	0.050	0.050	0.060	0.070	0.394	1.710
18-33-East Central Texas Plains	9	0.050	0.050	0.055	0.090	0.100	0.282	0.370
18-34-Western Gulf Coastal Plain	4	0.060	--	0.113	0.155	0.211	--	0.305
19-30-Edwards Plateau	6	0.020	--	0.032	0.055	0.435	--	1.780
19-31-Southern Texas Plains	1	--	--	--	0.050	--	--	0.050
19-32-Texas Blackland Prairies	44	0.060	0.060	0.060	0.110	0.320	0.879	2.205
19-33-East Central Texas Plains	18	0.163	0.247	0.527	0.733	0.867	0.899	1.195
19-34-Western Gulf Coastal Plain	2	0.602	--	--	0.626	--	--	0.650
20-33-East Central Texas Plains	0	--	--	--	--	--	--	0.000
20-34- Western Gulf Coastal Plain	2	0.060	--	--	0.650	--	--	1.240
21-30-Edwards Plateau	6	0.050	0.050	0.050	0.050	0.054	0.058	0.060
21-31-Southern Texas Plains	24	0.002	0.060	0.060	0.078	0.131	0.163	0.323
21-33-East Central Texas Plains	1	--	--	--	0.140	--	--	--
21-34-Western Gulf Coastal Plain	2	0.139	--	--	0.143	--	--	0.148
22-34-Western Gulf Coastal Plain	8	0.092	0.154	0.545	0.718	0.815	1.111	1.415
23-24-Chihuahuan Deserts	30	0.005	0.050	0.060	0.095	0.344	0.613	0.790
23-30-Edwards Plateau	4	0.060	--	0.060	0.060	0.060	--	0.060
23-31-Southern Texas Plains	22	0.040	0.051	0.063	0.100	0.140	0.217	0.248
23-34-Western Gulf Coastal Plain	7	0.074	0.078	0.085	0.228	0.248	0.261	0.270
24-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	7	0.100	0.100	0.115	0.140	0.403	0.645	0.660

Total Nitrogen (TN; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26-Southwestern Tablelands	7	0.40	0.45	0.51	0.71	1.56	2.79	4.19
2-26-Southwestern Tablelands	4	0.72	--	0.77	2.95	5.37	--	6.18
2-27-Central Great Plains	12	0.98	1.09	1.31	1.46	2.05	3.77	5.28
2-29-Cross Timbers	1	--	--	--	1.55	--	--	--
2-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
2-33-East Central Texas Plains	0	--	--	--	--	--	--	--
2-35-South Central Plains	4	0.81	--	0.81	0.87	0.99	--	1.15
3-32-Texas Blackland Prairies	4	0.85	--	0.88	0.92	2.35	--	6.57
3-33-East Central Texas Plains	6	0.91	--	1.11	1.45	2.92	--	10.95
3-35-South Central Plains	5	1.05	--	1.14	1.21	1.22	--	9.42
4-33-East Central Texas Plains	0	--	--	--	--	--	--	--
4-35-South Central Plains	27	0.59	0.64	0.72	0.90	1.06	1.21	6.46
5-32-Texas Blackland Prairies	1	--	--	--	1.44	--	--	--
5-33-East Central Texas Plains	0	--	--	--	--	--	--	--
5-34-Western Gulf Coastal Plain	1	--	--	--	0.80	--	--	--
5-35-South Central Plains	20	0.78	0.84	0.95	1.15	1.26	1.60	1.64
6-33-East Central Texas Plains	0	--	--	--	--	--	--	--
6-35-South Central Plains	13	0.56	0.64	0.80	0.83	0.93	0.97	1.24
7-34-Western Gulf Coastal Plain	3	1.24	--	--	1.63	--	--	2.23
8-29-Cross Timbers	0	--	--	--	--	--	--	--
8-30 Edwards Plateau	47	0.59	0.74	0.83	1.14	1.61	8.20	12.69
8-32-Texas Blackland Prairies	7	0.70	2.70	5.05	6.15	7.79	8.39	8.94
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--

8-34 West Gulf Plain	2	0.99	--	--	1.03	--	--	1.07
8-35-South Central Plains	13	0.47	0.81	0.99	1.14	1.22	1.80	13.70
9-34-Western Gulf Coastal Plain	1	--	--	--	1.32	--	--	--
10-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
10-34- Western Gulf Coastal Plain	11	2.89	4.54	5.23	5.62	5.96	6.26	7.02
10-35-South Central Plains	4	0.62	--	0.73	0.84	0.98	--	1.20
11-34-Western Gulf Coastal Plain	5	0.59	--	0.94	1.54	1.71	--	3.44
12-25-High Plains	2	3.18	--	--	3.72	--	--	4.25
12-26-Southwestern Tablelands	3	0.46	--	--	0.62	--	--	5.74
12-27-Central Great Plains	9	0.90	1.14	1.39	1.45	4.10	6.43	11.60
12-29-Cross Timbers	39	0.75	0.85	1.12	1.51	2.33	5.43	24.95
12-30-Edwards Plateau	6	0.37	--	0.47	0.66	0.79	--	2.03
12-32-Texas Blackland Prairies	18	0.97	1.13	1.84	2.63	3.93	6.88	8.55
12-33-East Central Texas Plains	13	0.64	0.85	1.20	1.91	11.10	32.52	82.50
12-34-Western Gulf Coastal Plain	3	1.35	--	--	1.38	--	--	5.34
13-34-Western Gulf Coastal Plain	8	1.12	1.22	1.28	1.40	1.51	1.55	1.62
14-25-High Plains	0	--	--	--	--	--	--	--
14-26-Southwestern Tablelands	2	1.34	--	--	1.40	--	--	1.46
14-27-Central Great Plains	7	0.89	1.17	1.37	1.43	1.77	2.34	2.90
14-29-Cross Timbers	7	0.83	0.84	0.85	0.92	1.09	4.63	9.77
14-30-Edwards Plateau	41	0.24	0.28	0.31	0.44	0.69	1.10	6.98
14-32-Texas Blackland Prairies	13	0.58	0.59	0.62	1.31	2.74	5.52	7.35
14-33-East Central Texas Plains	8	0.56	0.56	0.59	1.80	2.33	2.42	2.62
14-34-Western Gulf Coastal Plain	5	1.14	--	1.71	1.77	1.86	--	1.95
15-34-Western Gulf Coastal Plain	1	1.44	1.44	1.44	1.44	1.44	1.44	1.44
16-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
16-33-East Central Texas Plains	0	--	--	--	--	--	--	--
16-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
18-30-Edwards Plateau	2	0.52	--	--	0.80	--	--	1.08
18-32-Texas Blackland Prairies	12	0.49	0.52	0.69	1.29	1.68	1.96	7.47
18-33-East Central Texas Plains	1	--	--	--	1.65	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	0.55	--	--	--
19-30-Edwards Plateau	4	0.34	--	0.54	0.81	1.06	--	1.18
19-31-Southern Texas Plains	1	--	--	--	0.53	--	--	--
19-32-Texas Blackland Prairies	44	0.83	1.14	1.51	2.15	3.74	7.08	9.00
19-33-East Central Texas Plains	18	1.72	1.95	6.51	7.39	9.39	10.07	10.40
19-34-Western Gulf Coastal Plain	1	--	--	--	5.36	--	--	--
20-33-East Central Texas Plains	0	--	--	--	--	--	--	--
20-34- Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
21-30-Edwards Plateau	5	0.46	--	0.46	0.55	0.66	--	0.84
21-31-Southern Texas Plains	11	0.89	0.91	1.20	3.63	6.89	8.01	16.30
21-33-East Central Texas Plains	1	--	--	--	1.14	--	--	--
21-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
22-34-Western Gulf Coastal Plain	5	1.49	--	1.95	5.09	6.64	--	8.06
23-24-Chihuahuan Deserts	25	0.56	0.89	1.00	1.21	2.47	3.13	3.68
23-30-Edwards Plateau	4	1.23	--	1.41	1.52	1.64	--	1.85
23-31-Southern Texas Plains	12	0.62	0.81	0.98	1.31	1.84	2.04	2.67
23-34-Western Gulf Coastal Plain	5	0.86	--	0.87	1.37	1.48	--	1.65
24-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--

Nitrite Plus Nitrate-Nitrogen (NO_X-N; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26-Southwestern Tablelands	15	0.04	0.04	0.06	0.25	2.01	4.63	6.86

2-26-Southwestern Tablelands	11	0.04	0.04	0.16	0.35	2.17	3.35	4.35
2-27-Central Great Plains	20	0.02	0.04	0.08	0.22	0.59	2.01	3.72
2-29-Cross Timbers	1	--	--	--	0.06	--	--	--
2-32-Texas Blackland Prairies	3	0.06	--	--	0.49	--	--	9.34
2-33-East Central Texas Plains	3	0.10	--	--	0.14	--	--	0.23
2-35-South Central Plains	7	0.04	0.04	0.06	0.10	0.10	0.15	0.24
3-32-Texas Blackland Prairies	4	0.04	--	0.10	0.13	1.35	--	5.01
3-33-East Central Texas Plains	6	0.14	--	0.21	0.37	1.63	--	8.64
3-35-South Central Plains	5	0.05	--	0.12	0.18	0.20	--	10.22
4-33-East Central Texas Plains	0	--	--	--	--	--	--	--
4-35-South Central Plains	28	0.04	0.04	0.05	0.09	0.28	0.42	5.81
5-32-Texas Blackland Prairies	1	--	--	--	0.15	--	--	--
5-33-East Central Texas Plains	1	--	--	--	0.17	--	--	--
5-34-Western Gulf Coastal Plain	3	0.06	--	--	0.09	--	--	0.09
5-35-South Central Plains	27	0.08	0.09	0.12	0.15	0.22	0.34	0.74
6-33-East Central Texas Plains	0	--	--	--	--	--	--	--
6-35-South Central Plains	67	0.04	0.05	0.06	0.18	0.44	0.89	8.94
7-34-Western Gulf Coastal Plain	6	0.05	--	0.08	0.15	0.28	--	0.76
8-29-Cross Timbers	0	--	--	--	--	--	--	--
8-30 Edwards Plateau	47	0.08	0.13	0.25	0.41	0.88	6.97	11.30
8-32-Texas Blackland Prairies	7	0.12	1.57	3.59	5.20	6.33	6.97	7.72
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--
8-34 West Gulf Plain	5	0.09	--	0.15	0.23	0.26	--	3.54
8-35-South Central Plains	14	0.04	0.13	0.20	0.28	0.42	1.29	13.05
9-34-Western Gulf Coastal Plain	2	0.04	0.07	0.10	0.17	0.23	0.27	0.30
10-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
10-34- Western Gulf Coastal Plain	18	0.04	0.04	0.50	3.60	4.39	4.62	4.95
10-35-South Central Plains	32	0.04	0.04	0.04	0.13	0.31	0.81	12.39
11-34-Western Gulf Coastal Plain	41	0.02	0.04	0.09	0.27	1.08	2.13	3.92
12-25-High Plains	2	0.76	--	--	1.69	--	--	2.61
12-26-Southwestern Tablelands	6	0.04	--	0.04	0.07	0.49	--	1.88
12-27-Central Great Plains	11	0.04	0.04	0.08	0.23	3.39	10.16	11.30
12-29-Cross Timbers	83	0.04	0.06	0.10	0.20	0.68	1.76	19.31
12-30-Edwards Plateau	8	0.04	0.10	0.14	0.23	0.63	0.96	1.28
12-32-Texas Blackland Prairies	33	0.10	0.21	0.62	1.12	1.87	3.30	7.73
12-33-East Central Texas Plains	32	0.04	0.10	0.15	0.24	0.55	9.54	70.60
12-34-Western Gulf Coastal Plain	14	0.11	0.21	0.29	0.54	0.82	2.36	6.72
13-34-Western Gulf Coastal Plain	10	0.04	0.04	0.22	0.24	0.37	0.46	0.47
14-25-High Plains	2	0.48	--	--	6.52	--	--	12.56
14-26-Southwestern Tablelands	14	0.02	0.02	0.02	0.02	0.06	0.09	5.50
14-27-Central Great Plains	24	0.02	0.06	0.13	0.27	1.73	3.09	6.60
14-29-Cross Timbers	7	0.02	0.06	0.10	0.14	0.28	4.02	9.44
14-30-Edwards Plateau	73	0.02	0.03	0.06	0.11	0.43	1.69	5.36
14-32-Texas Blackland Prairies	25	0.04	0.14	0.26	0.46	1.04	2.30	6.45
14-33-East Central Texas Plains	8	0.02	0.02	0.02	0.71	1.77	1.93	2.27
14-34-Western Gulf Coastal Plain	5	0.10	--	1.03	1.12	1.27	--	1.44
15-34-Western Gulf Coastal Plain	1	--	--	--	0.50	--	--	--
16-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
16-33-East Central Texas Plains	2	0.26	--	--	0.28	--	--	0.30
16-34-Western Gulf Coastal Plain	5	0.08	--	0.16	0.21	0.23	--	0.81
18-30-Edwards Plateau	22	0.07	0.13	0.18	0.29	0.52	0.64	0.78
18-32-Texas Blackland Prairies	21	0.26	0.28	0.50	0.85	1.53	6.96	11.70
18-33-East Central Texas Plains	7	0.03	0.12	0.41	0.66	0.91	1.20	1.34
18-34-Western Gulf Coastal Plain	4	0.05	--	0.42	0.56	0.69	--	1.05

19-30-Edwards Plateau	4	0.14	--	0.31	0.48	0.68	--	0.98
19-31-Southern Texas Plains	1	--	--	--	0.28	--	--	--
19-32-Texas Blackland Prairies	45	0.04	0.35	0.84	1.90	3.38	6.24	7.96
19-33-East Central Texas Plains	18	1.25	1.37	5.98	6.73	8.54	9.37	9.53
19-34-Western Gulf Coastal Plain	2	2.56	--	--	3.58	--	--	4.60
20-33-East Central Texas Plains	0	--	--	--	--	--	--	--
20-34- Western Gulf Coastal Plain	2	0.07	--	--	1.21	--	--	2.34
21-30-Edwards Plateau	6	0.21	--	0.24	0.38	0.57	--	0.62
21-31-Southern Texas Plains	24	0.02	0.02	0.10	0.63	2.82	6.62	15.95
21-33-East Central Texas Plains	1	--	--	--	0.06	--	--	--
21-34-Western Gulf Coastal Plain	2	0.03	--	--	0.05	--	--	0.07
22-34-Western Gulf Coastal Plain	8	0.15	0.20	2.70	3.92	4.39	5.37	5.64
23-24-Chihuahuan Deserts	29	0.04	0.04	0.06	0.34	0.71	0.95	1.73
23-30-Edwards Plateau	4	0.35	--	1.05	1.33	1.44	--	1.66
23-31-Southern Texas Plains	20	0.06	0.12	0.18	0.27	0.48	1.63	1.66
23-34-Western Gulf Coastal Plain	7	0.15	0.16	0.17	0.43	0.52	0.58	0.66
24-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	7	0.04	0.04	0.04	0.04	0.04	0.04	0.05

Ortho-Phosphate (PO_4-P; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
1-26-Southwestern Tablelands	16	0.020	0.040	0.040	0.043	0.071	0.135	0.360
2-26-Southwestern Tablelands	18	0.020	0.020	0.020	0.040	0.048	0.237	0.750
2-27-Central Great Plains	24	0.020	0.040	0.040	0.053	0.066	0.299	1.180
2-29-Cross Timbers	2	0.040	--	--	0.040	--	--	0.040
2-32-Texas Blackland Prairies	3	0.056	--	--	0.205	--	--	3.860
2-33-East Central Texas Plains	5	0.045	--	0.050	0.085	0.100	--	0.770
2-35-South Central Plains	7	0.040	0.052	0.060	0.060	0.065	0.094	0.130
3-32-Texas Blackland Prairies	4	0.040	--	0.055	0.060	0.226	--	0.725
3-33-East Central Texas Plains	4	0.055	--	0.059	0.100	0.270	--	0.660
3-35-South Central Plains	3	0.060	--	--	0.060	--	--	0.060
4-33-East Central Texas Plains	0	--	--	--	--	--	--	--
4-35-South Central Plains	26	0.010	0.010	0.013	0.055	0.060	0.060	0.920
5-32-Texas Blackland Prairies	1	--	--	--	0.120	--	--	--
5-33-East Central Texas Plains	1	--	--	--	0.050	--	--	--
5-34-Western Gulf Coastal Plain	3	0.040	--	--	0.040	--	--	0.105
5-35-South Central Plains	28	0.040	0.040	0.040	0.040	0.050	0.118	0.865
6-33-East Central Texas Plains	0	--	--	--	--	--	--	--
6-35-South Central Plains	64	0.010	0.025	0.040	0.050	0.070	0.147	2.775
7-34-Western Gulf Coastal Plain	5	0.030	--	0.060	0.060	0.070	--	0.206
8-29-Cross Timbers	1	--	--	--	0.115	--	--	--
8-30 Edwards Plateau	82	0.010	0.020	0.040	0.040	0.058	0.704	2.210
8-32-Texas Blackland Prairies	8	0.020	0.034	0.250	0.525	0.828	0.840	0.850
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--
8-34 West Gulf Plain	6	0.040	--	0.049	0.060	0.060	--	0.900
8-35-South Central Plains	22	0.008	0.020	0.020	0.040	0.060	0.661	1.495
9-34-Western Gulf Coastal Plain	2	0.050	--	--	0.065	--	--	0.080
10-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
10-34- Western Gulf Coastal Plain	113	0.040	0.080	0.275	0.895	1.410	1.946	3.480
10-35-South Central Plains	39	0.010	0.018	0.040	0.060	0.115	0.922	2.920
11-34-Western Gulf Coastal Plain	40	0.040	0.040	0.069	0.225	0.380	0.661	0.780
12-25-High Plains	2	0.060	--	--	0.060	--	--	0.060
12-26-Southwestern Tablelands	6	0.040	--	0.040	0.040	0.048	--	0.780
12-27-Central Great Plains	11	0.040	0.040	0.040	0.170	0.985	2.485	2.660

12-29-Cross Timbers	119	0.003	0.006	0.040	0.040	0.065	0.356	1.790
12-30-Edwards Plateau	9	0.040	0.040	0.040	0.040	0.045	0.073	0.125
12-32-Texas Blackland Prairies	37	0.040	0.040	0.040	0.050	0.090	0.398	1.410
12-33-East Central Texas Plains	39	0.040	0.040	0.040	0.040	0.098	2.294	8.010
12-34-Western Gulf Coastal Plain	18	0.040	0.040	0.055	0.153	0.408	1.157	1.690
13-34-Western Gulf Coastal Plain	11	0.040	0.070	0.082	0.101	0.120	0.130	0.250
14-25-High Plains	1	--	--	--	1.860	--	--	--
14-26-Southwestern Tablelands	6	0.020	--	0.037	0.040	0.055	--	0.083
14-27-Central Great Plains	21	0.040	0.040	0.040	0.040	0.227	0.280	0.310
14-29-Cross Timbers	7	0.027	0.035	0.040	0.060	0.065	0.518	1.190
14-30-Edwards Plateau	72	0.020	0.020	0.020	0.025	0.040	0.058	1.550
14-32-Texas Blackland Prairies	25	0.020	0.020	0.020	0.033	0.061	0.164	0.430
14-33-East Central Texas Plains	8	0.040	0.040	0.040	0.165	0.293	0.318	0.360
14-34-Western Gulf Coastal Plain	5	0.060	--	0.170	0.210	0.210	--	0.276
15-34-Western Gulf Coastal Plain	1	--	--	--	0.180	--	--	--
16-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
16-33-East Central Texas Plains	0	--	--	--	--	--	--	--
16-34-Western Gulf Coastal Plain	5	0.065	--	0.120	0.120	0.140	--	0.145
18-30-Edwards Plateau	3	0.040	--	--	0.040	--	--	0.060
18-32-Texas Blackland Prairies	12	0.020	0.021	0.030	0.040	0.040	0.040	0.745
18-33-East Central Texas Plains	1	--	--	--	0.060	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	0.060	--	--	--
19-30-Edwards Plateau	6	0.020	--	0.020	0.030	0.385	--	1.800
19-31-Southern Texas Plains	1	--	--	--	0.040	--	--	--
19-32-Texas Blackland Prairies	37	0.020	0.020	0.040	0.063	0.446	0.902	2.240
19-33-East Central Texas Plains	14	0.134	0.153	0.344	0.705	0.741	0.962	1.120
19-34-Western Gulf Coastal Plain	1	--	--	--	0.420	--	--	--
20-33-East Central Texas Plains	0	--	--	--	--	--	--	--
20-34- Western Gulf Coastal Plain	2	0.020	--	--	0.595	--	--	1.170
21-30-Edwards Plateau	5	0.040	--	0.040	0.040	0.060	--	0.060
21-31-Southern Texas Plains	19	0.030	0.040	0.040	0.040	0.060	0.062	0.205
21-33-East Central Texas Plains	1	--	--	--	0.060	--	--	--
21-34-Western Gulf Coastal Plain	2	0.130	--	--	0.135	--	--	0.140
22-34-Western Gulf Coastal Plain	8	0.040	0.061	0.291	0.392	0.498	0.718	0.970
23-24-Chihuahuan Deserts	27	0.006	0.015	0.040	0.060	0.073	0.403	0.570
23-30-Edwards Plateau	4	0.040	--	0.040	0.040	0.045	--	0.060
23-31-Southern Texas Plains	21	0.006	0.040	0.050	0.060	0.060	0.110	0.190
23-34-Western Gulf Coastal Plain	6	0.040	--	0.060	0.120	0.180	--	0.190
24-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	8	0.030	0.037	0.040	0.068	0.125	0.299	0.390

Fluorometric Chlorophyll-a (Chl-a; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
1-26-Southwestern Tablelands	5	3.00	--	3.28	4.78	7.06	--	18.9
2-26-Southwestern Tablelands	2	38.2	--	--	38.4	--	--	38.6
2-27-Central Great Plains	5	3.00	--	5.28	19.4	21.5	-	37.9
2-29-Cross Timbers	1	--	--	--	39.9	--	--	--
2-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
2-33-East Central Texas Plains	0	--	--	--	--	--	--	--
2-35-South Central Plains	1	--	--	--	38.9	--	--	--
3-32-Texas Blackland Prairies	2	7.02	--	--	7.88	--	--	8.73
3-33-East Central Texas Plains	6	3.00	--	3.00	3.09	3.34	--	7.10
3-35-South Central Plains	5	3.00	--	4.10	5.85	24.1	--	28.4
4-33-East Central Texas Plains	0	--	--	--	--	--	--	--

4-35-South Central Plains	13	3.00	3.00	3.00	3.06	4.66	4.98	5.12
5-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
5-33-East Central Texas Plains	0	--	--	--	--	--	--	--
5-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
5-35-South Central Plains	0	--	--	--	--	--	--	--
6-33-East Central Texas Plains	0	--	--	--	--	--	--	--
6-35-South Central Plains	13	3.00	3.00	3.00	8.39	12.0	21.9	41.3
7-34-Western Gulf Coastal Plain	2	16.5	--	--	28.1	--	--	39.6
8-29-Cross Timbers	0	--	--	--	--	--	--	--
8-30 Edwards Plateau	6	3.08	--	5.56	10.5	11.2	--	14.9
8-32-Texas Blackland Prairies	1	--	--	--	3.00	--	--	--
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--
8-34 West Gulf Plain	2	20.7	--	--	20.9	--	--	21.1
8-35-South Central Plains	6	3.00	--	4.34	7.22	10.6	--	14.0
9-34-Western Gulf Coastal Plain	1	--	--	--	3.00	--	--	--
10-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
10-34- Western Gulf Coastal Plain	11	3.00	3.56	4.69	6.94	8.69	9.52	10.1
10-35-South Central Plains	3	3.00	--	--	3.00	--	--	3.80
11-34-Western Gulf Coastal Plain	5	3.00	--	3.00	3.00	3.39	--	15.8
12-25-High Plains	2	54.8	--	--	55.3	--	--	55.9
12-26-Southwestern Tablelands	3	3.00	--	--	11.5	--	--	36.9
12-27-Central Great Plains	8	5.27	6.97	10.4	14.2	37.8	4	72.2
12-29-Cross Timbers	54	3.00	3.28	3.30	6.54	13.8	16.7	37.0
12-30-Edwards Plateau	6	3.00	--	3.00	3.15	3.30	--	3.30
12-32-Texas Blackland Prairies	20	3.00	3.00	3.23	4.56	7.29	21.4	25.2
12-33-East Central Texas Plains	15	3.10	3.27	3.30	4.69	9.41	11.8	50.6
12-34-Western Gulf Coastal Plain	5	5.00	--	6.61	7.75	17.2	--	27.2
13-34-Western Gulf Coastal Plain	7	0.730	0.936	1.86	3.00	3.42	4.36	5.71
14-25-High Plains	0	--	--	--	--	--	--	--
14-26-Southwestern Tablelands	1	--	--	--	26.6	--	--	--
14-27-Central Great Plains	6	8.72	--	13.4	24.7	29.4	--	68.8
14-29-Cross Timbers	6	5.00	--	7.01	8.90	11.0	--	20.1
14-30-Edwards Plateau	25	3.00	3.00	3.00	5.00	5.00	7.03	47.2
14-32-Texas Blackland Prairies	7	2.00	2.24	2.70	5.00	5.00	5.00	5.00
14-33-East Central Texas Plains	7	5.00	5.00	5.00	5.00	5.05	33.8	76.8
14-34-Western Gulf Coastal Plain	5	5.00	5.00	5.00	5.00	7.55	8.79	9.62
15-34-Western Gulf Coastal Plain	1	--	--	--	9.31	--	--	--
16-32-Texas Blackland Prairies	0	--	--	--	--	--	--	--
16-33-East Central Texas Plains	--	--	--	--	--	--	--	--
16-34-Western Gulf Coastal Plain	--	--	--	--	--	--	--	--
18-30-Edwards Plateau	2	3.00	--	--	3.00	--	--	3.00
18-32-Texas Blackland Prairies	6	3.00	--	3.00	3.00	3.00	--	3.64
18-33-East Central Texas Plains	1	--	--	--	3.12	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	4.50	--	--	--
19-30-Edwards Plateau	1	--	--	--	3.00	--	--	--
19-31-Southern Texas Plains	1	--	--	--	3.00	--	--	--
19-32-Texas Blackland Prairies	11	3.00	3.00	3.00	3.00	3.20	5.57	13.7
19-33-East Central Texas Plains	3	3.00	--	--	3.14	--	--	4.18
19-34-Western Gulf Coastal Plain	--	--	--	--	--	--	--	--
20-33-East Central Texas Plains	--	--	--	--	--	--	--	--
20-34- Western Gulf Coastal Plain	2	5.00	--	--	5.00	--	--	5.00
21-30-Edwards Plateau	4	2.00	--	2.75	3.00	3.00	--	3.00
21-31-Southern Texas Plains	17	3.00	3.00	3.00	6.40	9.06	14.4	24.0
21-33-East Central Texas Plains	1	--	--	--	31.8	--	--	--
21-34-Western Gulf Coastal Plain	2	5.00	--	--	5.50	--	--	6.00

22-34-Western Gulf Coastal Plain	5	20.1	--	33.3	37.9	76.4	--	100
23-24-Chihuahuan Deserts	17	3.00	4.99	8.39	14.9	21.9	31.4	56.7
23-30-Edwards Plateau	4	3.00	--	3.00	3.00	4.14	--	7.54
23-31-Southern Texas Plains	4	3.00	--	3.00	3.00	3.00	--	3.00
23-34-Western Gulf Coastal Plain	2	4.70	--	--	5.14	--	--	5.58
24-24-Chihuahuan Deserts	--	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	--	--	--	--	--	--	--	--

Table A1.6.2: Frequency Distribution of Median Nutrient and Chlorophyll-a Concentrations among Level IV Ecoregions in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	21	0.005	0.060	0.060	0.060	0.400	0.600	0.790
24b-Chihuahuan Desert Grssland	1	--	--	--	0.183	--	--	--
24c-Low Mountains and Bajadas	6	0.090	--	0.120	0.210	0.329	--	0.730
24e-Stockton Plateau	2	0.050	--	--	0.050	--	--	0.050
25i-Llano Estacado	3	0.145	--	--	0.230	--	--	2.230
26a-Canadian/Cimarron Breaks	13	0.060	0.061	0.080	0.085	0.120	0.274	0.508
26b-Flat Tablelands & Valleys	5	0.060	--	0.060	0.080	0.140	--	0.282
26c-Caprock Canyon/Bdlnd/Brk	21	0.020	0.040	0.050	0.060	0.120	1.010	1.130
26d-Semiarid Canadian Breaks	4	0.070	--	0.108	0.125	0.145	--	0.190
27h-Red Prairie	30	0.050	0.060	0.060	0.060	0.088	0.135	2.975
27i-Broken Red Plains	16	0.075	0.111	0.138	0.228	0.308	0.785	1.300
27j-Limestone Plains	10	0.060	0.060	0.060	0.063	0.378	0.548	1.250
29b-Eastern Cross Timbers	9	0.060	0.060	0.080	0.220	0.980	1.208	1.960
29c-Western Cross Timbers	29	0.060	0.060	0.060	0.140	0.280	0.860	1.300
29d-Grand Prairie	14	0.029	0.057	0.060	0.073	0.088	0.289	1.880
29e-Limestone Cut Plain	57	0.040	0.060	0.060	0.080	0.120	0.400	1.230
29f-Carbonate Cross Timbers	0	--	--	--	--	--	--	--
30a-Edwards Plateau Woodland	17	0.050	0.058	0.060	0.060	0.060	0.068	1.895
30b-Llano Uplift	6	0.050	--	0.060	0.060	0.060	--	0.060
30c-Balcones Canyonlands	63	0.007	0.020	0.022	0.050	0.053	0.060	1.780
30d-Semiarid Edwards Plateau	13	0.060	0.060	0.060	0.060	0.060	0.060	0.075
31a-Northern Nueces Allv Plns	12	0.002	0.050	0.058	0.060	0.060	0.060	0.070
31c-Texas-Tamaulipan Thrnsrcb	18	0.050	0.054	0.063	0.101	0.140	0.166	0.323
31d-Rio Grande Fldpln/Terrace	17	0.040	0.066	0.085	0.107	0.150	0.230	0.248
32a-Northern Blackland Prairie	194	0.020	0.050	0.060	0.065	0.163	0.637	4.200
32b-S Blackland/Fayette Prair	4	0.085	--	0.089	0.155	0.268	--	0.410
32c-Floodplains & Low Terrace	10	0.165	0.188	0.369	0.815	1.158	1.818	2.880
33a-Northern Post Oak Savanna	8	0.060	0.088	0.115	0.169	0.646	1.169	1.680
33b-Southern Post Oak Savanna	47	0.050	0.060	0.090	0.230	0.772	1.029	7.430
33c-San Antonio Prairie	4	0.080	--	0.089	0.151	1.072	--	3.660
33d-Northern Prairie Outliers	0	--	--	--	--	--	--	--
33f-Floodplains & Low Terrace	16	0.060	0.128	0.319	0.380	0.923	1.020	1.165
34a-N Humid Gulf Cstal Prair	178	0.060	0.121	0.180	0.650	1.060	1.613	3.280
34b-S Subhumid Glf Cstl Prair	7	0.060	0.079	0.115	0.180	0.648	0.892	1.240
34c-Floodplains & Low Terrace	21	0.060	0.131	0.185	0.250	0.364	0.602	0.780
34f-Lower Rio Grnd Allv Fldpl	13	0.074	0.082	0.228	0.270	0.720	0.936	1.415
34g- TX-LA Coastal Marshes	4	0.060	--	0.071	0.117	0.162	--	0.170
34h-Mid-Coast Barrier Islnds & Cstal Mrshs	3	0.100	--	--	0.130	--	--	0.305
35a-Tertiary Uplands	57	0.023	0.060	0.080	0.112	0.168	0.233	1.290
35b-Floodplains & Low Terrace	43	0.060	0.060	0.075	0.100	0.140	0.148	0.392

35c-Pleistocene Flvl Terraces	2	0.210	--	--	0.272	--	--	0.333
35e-Southern Tertiary Uplands	28	0.040	0.060	0.114	0.160	0.211	1.205	3.300
35f-Flatwoods	45	0.060	0.060	0.060	0.085	0.180	1.329	3.285
35g-Red River Bottomlands	6	0.087	--	0.103	0.113	0.119	--	0.150

Total Nitrogen (TN; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	18	0.56	0.91	1.01	1.25	2.74	3.35	3.68
24b-Chihuahuan Desert Grssland	1	--	--	--	1.18	--	--	--
24c-Low Mountains and Bajadas	4	0.77	--	1.45	1.71	2.05	--	2.96
24e-Stockton Plateau	2	0.94	--	--	0.97	--	--	1.01
25i-Llano Estacado	2	3.18	--	--	3.72	--	--	4.25
26a-Canadian/Cimarron Breaks	4	0.48	--	1.06	1.56	2.44	--	4.19
26b-Flat Tablelands & Valleys	3	0.46	--	--	0.62	--	--	1.46
26c-Caprock Canyon/Bdlnd/Brk	6	0.72	--	0.92	3.22	5.58	--	6.18
26d-Semiarid Canadian Breaks	2	0.40	--	--	0.56	--	--	0.71
27h-Red Prairie	14	0.98	1.36	1.43	1.58	2.67	4.02	11.60
27i-Broken Red Plains	8	1.08	1.16	1.20	1.36	1.95	3.83	5.28
27j-Limestone Plains	6	0.89	--	1.02	1.40	1.53	--	5.14
29b-Eastern Cross Timbers	8	0.99	1.00	1.03	2.14	6.46	10.74	13.70
29c-Western Cross Timbers	17	0.78	0.84	0.92	1.16	1.43	5.17	24.95
29d-Grand Prairie	9	0.47	0.69	0.88	1.18	1.89	1.95	1.97
29e-Limestone Cut Plain	26	0.80	0.85	1.10	1.38	2.28	3.52	5.74
29f-Carbonate Cross Timbers	0	--	--	--	--	--	--	--
30a-Edwards Plateau Woodland	17	0.31	0.34	0.44	0.62	0.91	1.10	6.98
30b-Llano Uplift	6	0.24	--	0.27	0.30	0.31	--	0.52
30c-Balcones Canyonlands	34	0.25	0.30	0.37	0.46	0.72	1.06	2.23
30d-Semiarid Edwards Plateau	5	1.23	--	1.45	1.47	1.58	--	1.85
31a-Northern Nueces Allv Plns	10	0.53	0.87	1.66	4.78	6.98	8.84	16.30
31c-Texas-Tamaulipan Thrnsrcb	7	0.62	0.78	0.93	1.14	1.82	1.85	1.86
31d-Rio Grande Fldpln/Terrace	7	0.79	0.90	1.00	1.03	1.82	2.30	2.67
32a-Northern Blackland Prairie	131	0.49	0.74	0.95	1.43	2.85	6.98	12.69
32b-S Blackland/Fayette Prair	0	--	--	--	--	--	--	--
32c-Floodplains & Low Terrace	8	1.43	1.70	2.51	4.80	7.89	10.10	12.68
33a-Northern Post Oak Savanna	4	0.70	--	0.97	1.13	3.63	--	10.95
33b-Southern Post Oak Savanna	33	0.57	1.05	1.65	6.98	9.50	10.55	82.50
33c-San Antonio Prairie	3	0.81	--	--	1.62	--	--	37.60
33d-Northern Prairie Outliers	0	--	--	--	--	--	--	--
33f-Floodplains & Low Terrace	15	0.56	1.04	1.81	2.62	6.10	7.83	8.94
34a-N Humid Gulf Cstal Prair	25	0.59	1.19	1.43	2.89	5.34	5.97	7.02
34b-S Subhumid Glf Cstl Prair	2	1.49	--	--	1.72	--	--	1.95
34c-Floodplains & Low Terrace	12	0.55	1.14	1.37	1.61	1.79	1.94	5.36
34f-Lower Rio Grnd Allv Fldpl	8	0.86	0.87	1.25	1.56	5.48	7.07	8.06
34g- TX-LA Coastal Marshes	3	0.80	--	--	1.24	--	--	2.23
34h-Mid-Coast Barrier Islnds & Cstal Mrshs	0	--	--	--	--	--	--	--
35a-Tertiary Uplands	27	0.53	0.64	0.77	0.95	1.07	1.23	6.46
35b-Floodplains & Low Terrace	36	0.63	0.71	0.81	1.00	1.21	1.45	1.64
35c-Pleistocene Flvl Terraces	1	--	--	--	9.42	--	--	--
35e-Southern Tertiary Uplands	2	0.62	--	--	0.91	--	--	1.20
35f-Flatwoods	4	0.56	--	0.72	0.81	0.86	--	0.90
35g-Red River Bottomlands	5	0.81	--	0.81	0.84	0.93	--	1.15

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	19	0.04	0.04	0.04	0.28	0.65	0.84	1.73
24b-Chihuahuan Desert Grssland	1	--	--	--	0.19	--	--	--

24c-Low Mountains and Bajadas	7	0.06	0.09	0.21	0.34	0.81	1.02	1.17
24e-Stockton Plateau	2	0.57	--	--	0.70	--	--	0.82
25i-Llano Estacado	4	0.48	--	0.69	1.69	5.10	--	12.56
26a-Canadian/Cimarron Breaks	12	0.04	0.04	0.04	0.14	1.49	4.11	6.86
26b-Flat Tablelands & Valleys	9	0.02	0.02	0.02	0.04	0.09	0.95	3.35
26c-Caprock Canyon/Bdlnd/Brk	20	0.02	0.02	0.03	0.08	0.83	3.21	5.50
26d-Semiarid Canadian Breaks	4	0.05	--	0.10	0.24	1.49	--	4.90
27h-Red Prairie	31	0.04	0.11	0.21	0.32	1.90	5.10	11.30
27i-Broken Red Plains	14	0.02	0.05	0.09	0.16	0.23	1.43	3.72
27j-Limestone Plains	10	0.02	0.03	0.04	0.05	0.76	2.59	4.67
29b-Eastern Cross Timbers	11	0.06	0.22	0.37	0.86	3.21	8.45	13.05
29c-Western Cross Timbers	33	0.02	0.05	0.09	0.15	0.41	1.04	19.31
29d-Grand Prairie	18	0.04	0.07	0.08	0.11	0.27	0.77	1.15
29e-Limestone Cut Plain	41	0.04	0.10	0.13	0.25	0.66	1.78	4.75
29f-Carbonate Cross Timbers	2	0.04	--	--	0.04	--	--	0.04
30a-Edwards Plateau Woodland	17	0.03	0.04	0.09	0.28	0.58	0.72	5.36
30b-Llano Uplift	6	0.02	--	0.02	0.02	0.03	--	0.05
30c-Balcones Canyonlands	81	0.03	0.07	0.10	0.21	0.49	0.78	3.06
30d-Semiarid Edwards Plateau	13	0.02	0.03	0.08	1.15	1.66	1.81	1.97
31a-Northern Nueces Allv Plns	12	0.04	0.35	0.98	2.99	6.38	7.28	15.95
31c-Texas-Tamaulipan Thrnsrcb	18	0.02	0.02	0.10	0.18	0.67	1.63	1.66
31d-Rio Grande Fldpln/Terrace	15	0.06	0.11	0.13	0.25	0.37	0.49	0.73
32a-Northern Blackland Prairie	162	0.04	0.14	0.34	0.80	2.05	6.09	11.70
32b-S Blackland/Fayette Prair	2	0.28	--	--	0.45	--	--	0.62
32c-Floodplains & Low Terrace	15	0.24	0.29	0.60	1.04	3.69	6.67	11.18
33a-Northern Post Oak Savanna	7	0.10	0.11	0.13	0.14	0.37	3.76	8.64
33b-Southern Post Oak Savanna	54	0.02	0.05	0.20	0.70	6.83	9.36	70.60
33c-San Antonio Prairie	4	0.10	--	0.14	0.20	4.50	--	17.27
33d-Northern Prairie Outliers	0	--	--	--	--	--	--	--
33f-Floodplains & Low Terrace	21	0.02	0.17	0.22	1.38	2.54	6.19	7.72
34a-N Humid Gulf Cstal Prair	85	0.02	0.04	0.06	0.25	1.65	3.71	4.95
34b-S Subhumid Glf Cstl Prair	5	0.03	--	0.07	0.15	0.23	--	2.34
34c-Floodplains & Low Terrace	28	0.04	0.10	0.23	0.43	1.06	2.48	6.72
34f-Lower Rio Grnd Allv Fldpl	13	0.15	0.16	0.43	0.66	3.99	5.03	5.64
34g- TX-LA Coastal Marshes	4	0.08	--	0.09	0.09	0.26	--	0.76
34h-Mid-Coast Barrier Islnds & Cstal Mrshs	3	0.04	--	--	0.05	--	--	1.05
35a-Tertiary Uplands	56	0.04	0.05	0.10	0.24	0.43	0.83	5.81
35b-Floodplains & Low Terrace	43	0.04	0.05	0.07	0.12	0.20	0.25	0.67
35c-Pleistocene Flvl Terraces	2	0.04	--	--	5.13	--	--	10.22
35e-Southern Tertiary Uplands	25	0.04	0.04	0.04	0.17	0.50	1.25	8.94
35f-Flatwoods	38	0.04	0.04	0.05	0.11	0.22	0.57	12.39
35g-Red River Bottomlands	6	0.04	--	0.09	0.10	0.10	--	0.24

Phosphate (PO_4-P; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	18	0.006	0.007	0.040	0.043	0.079	0.396	0.550
24b-Chihuahuan Desert Grssland	1	--	--	--	0.060	--	--	--
24c-Low Mountains and Bajadas	6	0.020	--	0.049	0.060	0.068	--	0.570
24e-Stockton Plateau	2	0.040	--	--	0.050	--	--	0.060
25i-Llano Estacado	3	0.060	--	--	0.060	--	--	1.860
26a-Canadian/Cimarron Breaks	13	0.040	0.040	0.040	0.040	0.060	0.150	0.360
26b-Flat Tablelands & Valleys	6	0.040	--	0.040	0.040	0.055	--	0.083
26c-Caprock Canyon/Bdlnd/Brk	22	0.020	0.020	0.020	0.040	0.050	0.471	0.780
26d-Semiarid Canadian Breaks	4	0.020	--	0.035	0.050	0.071	--	0.105

27h-Red Prairie	30	0.020	0.040	0.040	0.060	0.220	0.300	2.660
27i-Broken Red Plains	17	0.020	0.040	0.040	0.040	0.082	0.545	1.180
27j-Limestone Plains	9	0.040	0.040	0.040	0.040	0.270	0.574	1.670
29b-Eastern Cross Timbers	12	0.020	0.040	0.040	0.160	0.550	0.994	1.790
29c-Western Cross Timbers	42	0.027	0.040	0.040	0.050	0.070	0.413	1.190
29d-Grand Prairie	29	0.008	0.020	0.020	0.040	0.040	0.082	1.495
29e-Limestone Cut Plain	65	0.003	0.005	0.008	0.040	0.060	0.304	1.290
29f-Carbonate Cross Timbers	2	0.040	--	--	0.040	--	--	0.040
30a-Edwards Plateau Woodland	17	0.025	0.040	0.040	0.040	0.040	0.060	1.550
30b-Llano Uplift	6	0.030	--	0.040	0.040	0.040	--	0.040
30c-Balcones Canyonlands	67	0.020	0.020	0.020	0.020	0.040	0.051	1.800
30d-Semiarid Edwards Plateau	9	0.040	0.040	0.040	0.060	0.172	0.224	0.237
31a-Northern Nueces Allv Plns	10	0.040	0.040	0.040	0.040	0.058	0.060	0.060
31c-Texas-Tamaulipan Thrnsrcb	15	0.030	0.040	0.040	0.050	0.060	0.066	0.205
31d-Rio Grande Fldpln/Terrace	16	0.006	0.043	0.058	0.060	0.091	0.135	0.190
32a-Northern Blackland Prairie	185	0.010	0.020	0.040	0.040	0.066	0.469	3.860
32b-S Blackland/Fayette Prair	0	--	--	--	--	--	--	--
32c-Floodplains & Low Terrace	17	0.020	0.040	0.050	0.227	0.809	1.478	2.210
33a-Northern Post Oak Savanna	7	0.040	0.043	0.048	0.060	0.093	0.368	0.770
33b-Southern Post Oak Savanna	48	0.020	0.040	0.040	0.068	0.705	1.465	8.010
33c-San Antonio Prairie	4	0.040	--	0.040	0.045	0.986	--	3.795
33d-Northern Prairie Outliers	0	--	--	--	--	--	--	--
33f-Floodplains & Low Terrace	23	0.040	0.040	0.040	0.270	0.400	0.792	0.850
34a-N Humid Gulf Cstal Prair	179	0.030	0.055	0.090	0.520	1.100	1.812	3.480
34b-S Subhumid Glf Cstl Prair	7	0.020	0.032	0.055	0.140	0.325	0.702	1.170
34c-Floodplains & Low Terrace	29	0.040	0.048	0.101	0.160	0.260	0.461	1.685
34f-Lower Rio Grnd Allv Fldpl	12	0.040	0.060	0.150	0.277	0.412	0.595	0.970
34g- TX-LA Coastal Marshes	4	0.040	--	0.040	0.050	0.060	--	0.060
34h-Mid-Coast Barrier Islnds & Cstal Mrshs	1	--	--	--	0.080	--	--	--
35a-Tertiary Uplands	56	0.010	0.025	0.040	0.050	0.060	0.173	0.920
35b-Floodplains & Low Terrace	42	0.010	0.020	0.040	0.040	0.060	0.060	0.150
35c-Pleistocene Flvl Terraces	1	--	--	--	0.130	--	--	--
35e-Southern Tertiary Uplands	30	0.010	0.019	0.040	0.060	0.116	0.911	2.775
35f-Flatwoods	38	0.010	0.020	0.030	0.040	0.108	0.692	2.920
35g-Red River Bottomlands	5	0.060	--	0.060	0.060	0.060	--	0.070

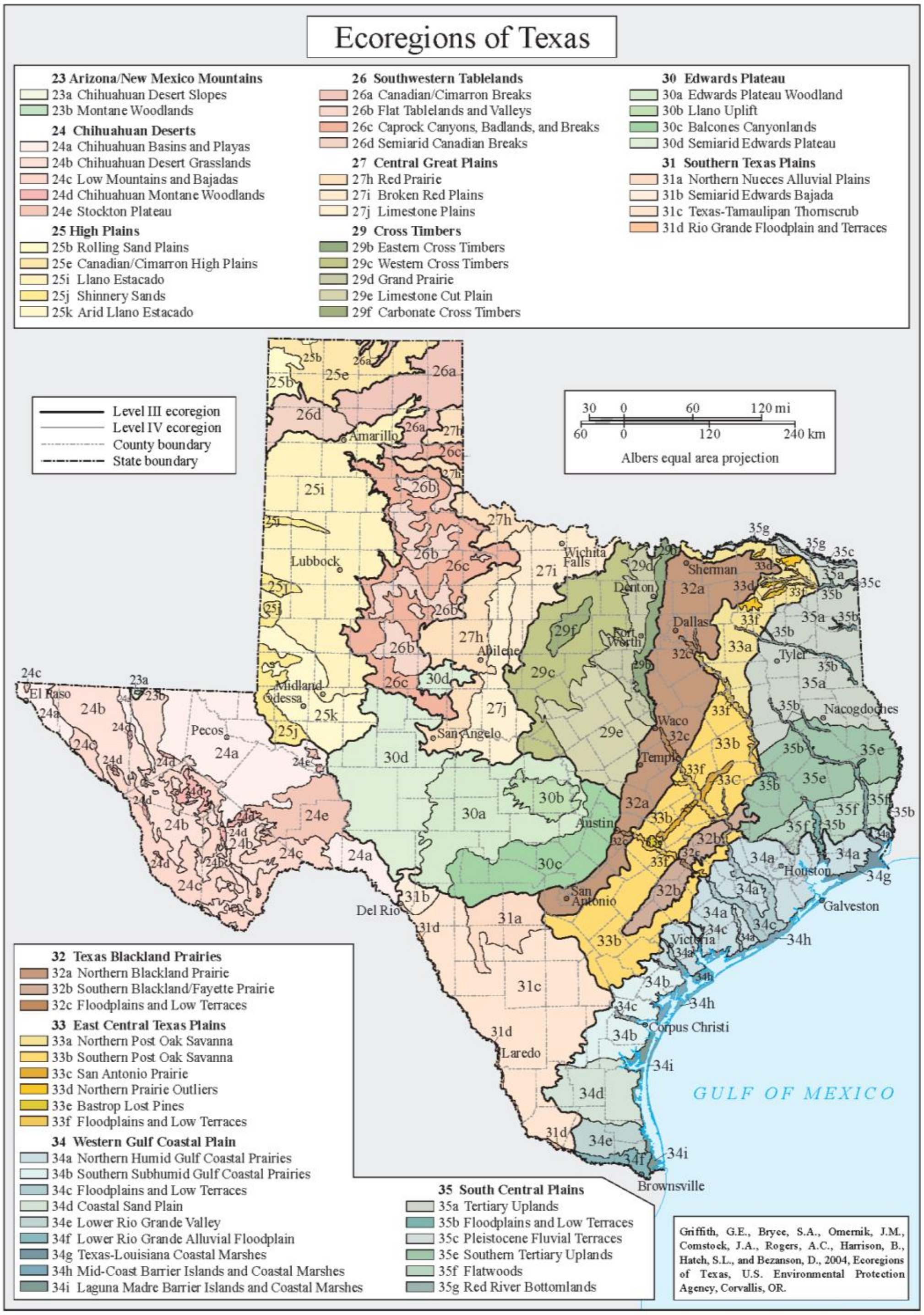
Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	10	5.70	7.50	8.95	14.4	20.4	22.7	29.7
24b-Chihuahuan Desert Grssland	1	--	--	--	21.1	--	--	--
24c-Low Mountains and Bajadas	4	12.2	--	20.8	28.8	39.6	--	56.7
24e-Stockton Plateau	2	3.00	--	--	3.46	--	--	3.92
25i-Llano Estacado	2	54.8	--	--	55.3	--	--	55.9
26a-Canadian/Cimarron Breaks	3	4.78	--	--	7.06	--	--	18.9
26b-Flat Tablelands & Valleys	3	3.00	--	--	11.5	--	--	26.6
26c-Caprock Canyon/Bdlnd/Brk	3	36.9	--	--	38.2	--	--	38.6
26d-Semiarid Canadian Breaks	1	--	--	--	3.28	--	--	--
27h-Red Prairie	11	3.00	5.27	6.49	13.9	29.2	68.8	72.2
27i-Broken Red Plains	4	14.5	--	18.2	20.4	25.6	--	37.9
27j-Limestone Plains	4	8.72	--	10.7	24.4	37.8	--	38.7
29b-Eastern Cross Timbers	6	3.28	--	7.33	11.2	14.5	--	39.9
29c-Western Cross Timbers	24	3.52	3.90	5.00	8.95	15.9	20.3	37.0
29d-Grand Prairie	5	3.00	--	9.85	11.6	14.0	--	14.0
29e-Limestone Cut Plain	32	3.00	3.00	3.30	3.30	8.41	14.4	21.5
29f-Carbonate Cross Timbers	0	--	--	--	--	--	--	--

30a-Edwards Plateau Woodland	15	3.00	3.00	3.00	5.00	5.92	9.06	47.2
30b-Llano Uplift	5	3.00	--	5.00	5.00	5.00	--	5.00
30c-Balcones Canyonlands	18	2.00	3.00	3.00	3.00	3.00	3.30	5.00
30d-Semiarid Edwards Plateau	4	3.00	--	3.00	3.00	4.14	--	7.54
31a-Northern Nueces Allv Plns	8	3.00	3.00	3.00	3.00	4.60	11.5	20.4
31c-Texas-Tamaulipan Thrnsrcb	12	3.00	3.00	4.50	6.48	9.37	10.4	24.0
31d-Rio Grande Fldpln/Terrace	2	3.00	--	--	3.00	--	--	3.00
32a-Northern Blackland Prairie	45	2.00	3.00	3.00	3.30	5.00	9.91	21.7
32b-S Blackland/Fayette Prair	0	--	--	--	--	--	--	--
32c-Floodplains & Low Terrace	7	2.40	3.96	5.00	10.2	16.37	22.9	25.2
33a-Northern Post Oak Savanna	3	3.00	--	--	3.00	--	--	3.00
33b-Southern Post Oak Savanna	19	3.00	3.11	3.28	5.00	10.1	35.5	76.8
33c-San Antonio Prairie	3	3.90	--	--	3.94	--	--	4.69
33d-Northern Prairie Outliers	0	--	--	--	--	--	--	--
33f-Floodplains & Low Terrace	9	3.00	3.14	3.40	5.00	5.00	5.50	7.10
34a-N Humid Gulf Cstal Prair	24	0.730	3.00	3.00	5.08	8.54	9.59	39.6
34b-S Subhumid Glf Cstl Prair	5	5.00	--	5.00	5.00	76.4	--	100
34c-Floodplains & Low Terrace	14	2.64	3.14	4.63	5.50	7.70	16.8	27.2
34f-Lower Rio Grnd Allv Fldpl	5	4.70	--	5.58	20.1	33.3	--	37.9
34g- TX-LA Coastal Marshes	1	--	--	--	16.5	--	--	--
34h-Mid-Coast Barrier Islnds & Cstal Mrshs	0	--	--	--	--	--	--	--
35a-Tertiary Uplands	16	3.00	3.00	3.00	3.03	4.78	21.5	41.3
35b-Floodplains & Low Terrace	16	3.00	3.00	4.33	7.26	14.16	22.6	28.4
35c-Pleistocene Flvl Terraces	1	--	--	--	4.10	--	--	--
35e-Southern Tertiary Uplands	1	--	--	--	3.80	--	--	--
35f-Flatwoods	3	3.00	--	--	3.00	--	--	3.00
35g-Red River Bottomlands	1	--	--	--	38.9	--	--	--

Appendix 1.7. Texas River and Coastal Basins (http://www.tceq.texas.gov/publications/gi/gi-316/gi-316_intro.html/at_download/file).

Appendix 1.8. Level III and Level IV Ecoregions in Texas (epa.gov/wed/ecoregions/tx/tx_eco_pg.pdf).



Appendix 1.9. Guide to interpreting CART and nCPA models and summary statistics

This appendix is intended as a guide to the users of this report for interpreting Classification and Regression Tree (CART) and non-parametric changepoint analysis (nCPA) stressor-response models, including thresholds, confidence intervals, r^2, and p values. For example, model summary statistics such as r^2, p values, and confidence intervals can assist the user in assessing the strength of the changepoint relationship identified by regression tree and non-parametric changepoint procedures. In addition to the threshold value itself, CART models also provide information on how many observations are grouped above and below the changepoint, as well as the average value of the response variable above and below the changepoint. In this user's guide to CART and nCPA models, we will 1) define the model summary statistics listed above, 2) describe and provide example uses for information appearing on the graphical display of a changepoint model, and 3) explore analyses on datasets generated to present idealized model outcomes to determine how summary statistics can vary with trends in the data.

Interpreting terminology of model output and summary statistics

Changepoint – The model changepoint is the value of the stressor variable that maximizes differences in the response variable related to magnitude and variability. If we assume that the data included in the analysis represent the general population without bias, we can assume the changepoint applies to the general population. Both CART and nCPA produce changepoint estimates. Usually these estimates are identical; however, in cases where r^2 differed between CART and nCPA, nCPA output was reported.

Confidence interval – The model confidence interval describes a range of values in the stressor variable surrounding the changepoint and bounded by the lower 5% confidence estimate and the upper 95% confidence estimate. Assuming that the data included in the analysis represent the general population without bias, we can be 95% confident that the changepoint value for the population falls within this range of values. Therefore, a more narrow confidence interval usually indicates a relatively small level of uncertainty associated with the changepoint estimate, also indicating a relatively strong model. However, it should be noted that confidence intervals may not accurately reflect uncertainty associated with the changepoint if bias exists in the data, such as censoring at a common detection limit. Only nCPA provides estimates of confidence levels.

R^2 - The model r^2 describes the variability in the data that can be explained by the model. Therefore, the higher the r^2, the greater the explanatory power and strength of the model. For example, if r^2=0.25, the identified changepoint in the stressor variable describes 25% of the variability in the response variable. Both CART and nCPA produce r^2 estimates. Usually these estimates are identical; however, in cases where model r^2 differed between CART and nCPA, output from nCPA was reported.

p_{perm}- The model p_{perm} is a measure of the statistical probability that a changepoint relationship exists between the stressor and response variables at the identified threshold level in the stressor variable. The acceptable level of uncertainty was defined for these analyses to be $p_{perm} < 0.05$. In other words, only results with a less than 5% probability of error were accepted as statistically significant. Therefore, the smaller the p_{perm}, the greater the probability that the model describes relationships in the general population, assuming that the data included in the analysis represent the general population without bias. Only nCPA provides estimates of p_{perm}.

Interpreting figures illustrating CART and nCPA models

The standard template for figures representing changepoint and regression tree models that was used throughout this report is shown in Figure A1.10.1. The changepoint model developed to describe the relationship between median chlorophyll-a measured spectrophotometrically (chl-a spec) and median total nitrogen (TN) concentrations in Texas reservoirs was randomly selected for this illustration (medians are from Dataset 1, see Sections 2.3-2.4). Notes have been added to the standard template for figures in order to assist with interpretation of graphical representations of regression tree and changepoint analysis results.

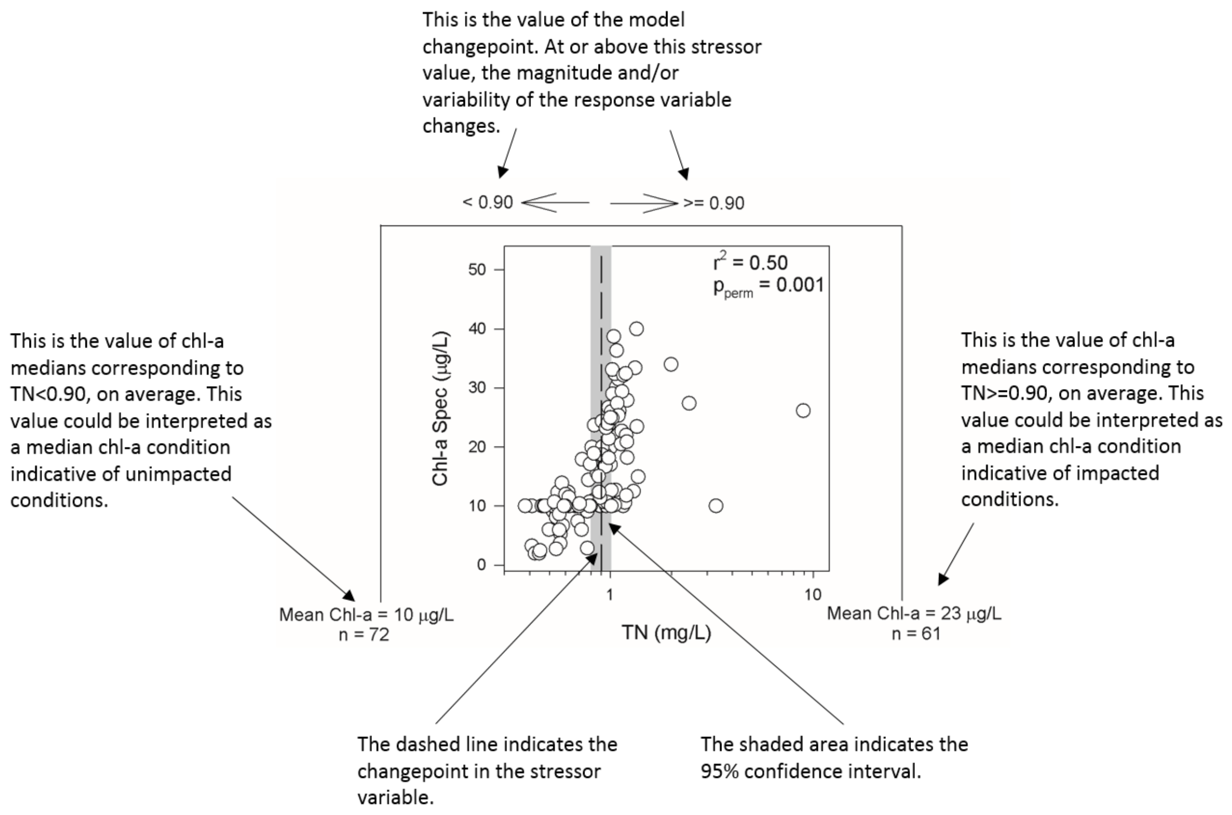


Figure A1.10.1. CART model of median spectrophotometric chl-a (chl-a spec) vs. medians TN concentrations for Texas reservoirs. Medians were from Dataset 1 (see Sections 2.3-2.4).

The water quality data included in CART and nCPA analysis were consistently station medians or means. Raw data were never analyzed. Therefore, the scatterplots representing water quality data throughout this report always depict measures of central tendency for waterbody stations. Where CART and nCPA yielded a statistically significant model, the value of the threshold is listed above the scatterplot. For statistically significant models, the changepoint is shown as a dashed line and the confidence interval is shown as a shaded area surrounding the changepoint. The number of observations and average median or mean value of the response variable associated with stressor values above or below the threshold are include to the right and left of the scatterplot, respectively.

Example scenarios

In order to explore how model summary statistics vary with specific datasets, 4 datasets were create to generate idealized changepoint models. These data were not derived from the TCEQ water quality databases and do not specifically represent any trends that may be present in those data.

Scenario 1. This data scenario is illustrated in Figure A1.x.2. The analysis identified a changepoint in response to the stressor variable equal to 50.5. The dataset used to generate this model represents a "perfect" changepoint relationship between the stressor and response variables. Every value in the response data that is paired with a stressor value <50.5 is equal to 1, while every response value paired with a stressor value ≥50.5 is equal to 2. In this scenario, a model that defines a changepoint in the stressor variable = 50.5 explains all variability in the response data. Therefore, the model r^2=1.

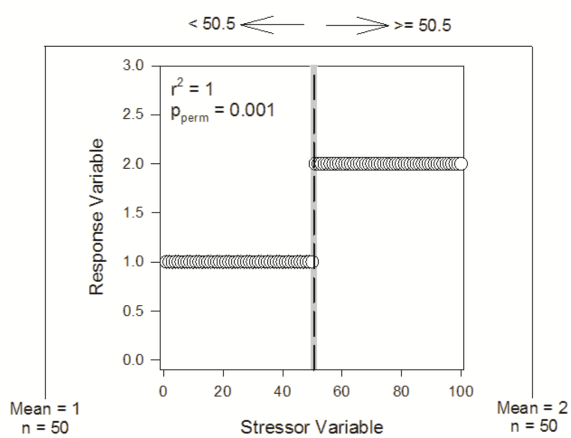


Figure A1.10.2. Sample model describing the changepoint relationship between a stressor and a response variable. In this scenario, the data grouped above and below the threshold are always different between groups, but no variability is present among groups.

In this scenario, the model identifying a changepoint in the stressor variable = 50.5 also has a low probability of error, as indicated by a $p_{perm} \leq 0.001$. This p_{perm} is < 0.05, the criteria required for the model to be statistically significant. Similarly, the confidence interval surrounding the threshold estimate is approximately 2% of the total stressor data range. Therefore, we can be 95% confident the changepoint is located within 49.5 to 51.5. Indeed the confidence interval, represented by the shaded area surrounding the threshold line, is barely visible.

In summary, we can conclude that this CART model provides excellent explanatory power for the dataset, and that it is highly probable that the analysis has identified a real relationship between the stressor and response variables. The narrow confidence interval indicates that the range of possible alternative threshold values is small. Scenario 1 is a highly idealized representation of a threshold relationship between two variables. It is highly improbable that this scenario would arise from analysis of an environmental dataset.

Scenario 2. This data scenario is illustrated in Figure A1.x.3. As in Scenario 1, the analysis identified a changepoint in response to the stressor variable = 50.5. However, the dataset used to generate this model differs from the dataset in Scenario 1. While values in the response variable corresponding to stressor values above and below the threshold never overlap, variability is now present in the data grouped above and below the threshold. Every value in the response data that is paired with a stressor value <50.5 falls within a range of values between 0 and 1, while every response value paired with a stressor value ≥50.5 is between 1.5 and 2.5. For this model, r^2 =0.77, a 23% reduction from Scenario 1.

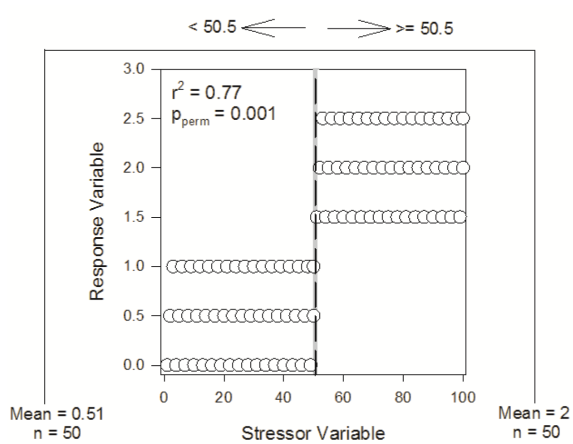


Figure A1.x.3. Sample model describing the changepoint relationship between a stressor and a response variable. In this scenario, the data grouped above and below the threshold are always different between groups, but variability is present among groups.

In this scenario, the model identifying a changepoint in the stressor variable = 50.5 has a low probability of error, as indicated by a $p_{perm} \leq 0.001$. This p_{perm} is < 0.05, the criteria required for the model to be statistically significant. Similarly, the confidence interval surrounding the threshold estimate is 1.5% of the total stressor data range. Therefore, we can be 95% confident the changepoint is within 50 to 51.5.

As in Scenario 1, we can conclude that the model provides excellent explanatory power, though not 100%. It is also highly probable that the analysis has identified a real relationship between the stressor and response variables, and the range of possible alternative threshold values is small. Scenario 2 is also an idealized representation of a threshold relationship between two variables, and achieving model explanatory power of 77% is unlikely in environmental datasets. However, when the "real-world" model in Fig. A1.10.1. is compared with Scenario 2, it is apparent that the variability in the data relative to the threshold is similar, though the ranges of values above and below the threshold overlap somewhat in Fig. A1.10.1. Despite this similarity and the obvious threshold relationship in Fig. A1.10.1, explanatory power for the model in Fig. A1.10.1 is only 50%, illustrating the magnitude of r^2 values that are likely in changepoint analysis of environmental data. In fact, r^2=0.50 is among the highest observed in this study.

Scenario 3. This data scenario is illustrated in Figure A1.10.4. The analysis identified a changepoint in response to the stressor variable equal to 51.5. In this dataset, values in the response variable corresponding to stressor values above and below the threshold overlap and variability is present in the data above and below the threshold. Every value in the response data paired with a stressor value <51.5 falls within a range of values between 0 and 1, while every response value paired with a stressor value ≥51.5 is between 0.5 and 2. In other words, mid-range response variable data appear on both sides of the threshold, but the lowest and highest values only appear below or above the threshold, respectively. For this model, r^2 =0.38, a 62% reduction from Scenario 1 and a 50% reduction from Scenario 2.

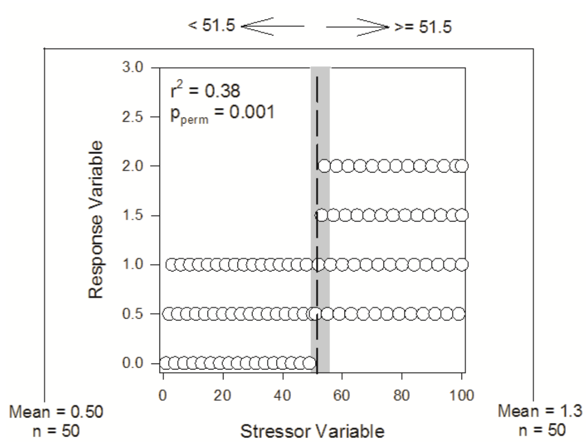


Figure A1.10.4. Sample model describing the changepoint relationship between a stressor and a response variable. In this scenario, the data grouped above and below the threshold overlap in the mid-range of observed values and variability is present among groups. Only the highest and lowest values are unique to the right or left of the threshold, respectively.

In this scenario, the model identifying a changepoint in the stressor variable equal to 51.5 also has a low probability of error, as indicated by a $p_{perm} \le 0.001$. This p_{perm} is < 0.05, the criteria required for the model to be statistically significant. Similarly, the confidence interval surrounding the threshold estimate is approximately 6.5% of the total stressor data range. Therefore, we can be 95% confident the changepoint is located within 49.5 to 56. Though still a narrow range, the confidence interval for the model in Scenario 3 is approximately 3-5x greater than for the models in Scenarios 1-2.

We can conclude that this model provides good explanatory power. It is also highly probable that the analysis has identified a real relationship between the stressor and response variables, and the range of possible alternative threshold values is small. Scenario 3 is a less idealized representation of a threshold relationship between two variables than Scenarios 1-2, and many of the relationships in the TCEQ water quality databases were similar to that shown in Fig. A1.x.4 and had similar r^2. It is key to note, however, that the "real-world" water quality models that had similar explanatory power to the example in Scenario 3 were among the strongest models identified in the water quality database.

Scenario 4. This data scenario is illustrated in Figure A1.x.5. In this dataset, the analysis identified a changepoint in response to the stressor variable equal to 52.5.In this dataset, values in the response variable corresponding to stressor values above and below the threshold overlap and variability is present in the data above and below the threshold. Every value in the response data that is paired with a stressor value <52.5 falls within a range of values between 0 and 1, while every response value paired with a stressor value ≥52.5 is between 0 and 2.5. In other words, all response variable values associated with stressor values below the threshold are also present above the threshold. However, the highest values in the dataset are only present above the threshold. For this model r^2 =0.29, a 70% reduction from Scenario 1 and a 60% reduction from Scenario 2.

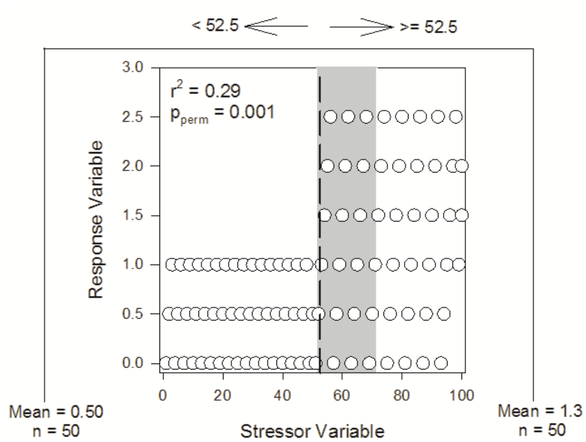


Figure A1.x.5. Sample model describing the changepoint relationship between a stressor and a response variable. In this scenario, the data grouped above and below the threshold overlap in value in the mid-range of observed values and variability is present among groups. Only the highest and lowest values are unique to the right or left of the threshold, respectively.

In this scenario, the model identifying a changepoint in the stressor variable equal to 52.5 also has a low probability of error, as indicated by a $p_{perm} \leq 0.001$. This p_{perm} is < 0.05, the criteria required for the model to be statistically significant. In contrast, the confidence interval surrounding the threshold estimate is much wider for Scenario 4 than for any of the previous scenarios, approximately 25% of the total stressor data range. While the analysis identified 52.5 as the changepoint, the value of the threshold could be as low as 51.5 and as high as 75.

We can conclude that this model provides good explanatory power. It is also highly probable that the analysis has identified a real relationship between the stressor and response variables. However, in contrast to previous scenarios, the 95% confidence interval is 25% of the range of possible values. Scenario 4 is a less idealized representation of a threshold relationship between two variables than Scenarios 1-2, and many of the relationships in the TCEQ water quality databases were similar to that shown in Fig. A1.x.5 and had similar r^2. It is key to note, however, that the "real-world" water quality models that had similar explanatory power to the example in Scenario 4 were among the strongest models identified in the water quality database. In the TCEQ data, similar trends to that shown in A1.x.5 are often visible, but analysis may also indicate an r^2 as low as 0.05.

Appendix 1.10. Classification and regression tree (CART) and non-parametric changepoint analysis code from geospatial analyses

ANALYSIS: TP VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TP ~ DEV + DEVAG + AG + FOR + WET, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=1060 (1130 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.12255228      0 1.0000000 1.0026733 0.1685396
2 0.04815059      1 0.8774477 0.8950497 0.1538030

Node number 1: 1060 observations,    complexity param=0.1225523
  mean=0.3497025, MSE=0.3364858
  left son=2 (637 obs) right son=3 (423 obs)
  Primary splits:
      DEV   < 0.1117865  to the left,  improve=0.12255230, (0 missing)
      DEVAG < 0.5399513  to the left,  improve=0.09041503, (0 missing)
      FOR   < 0.3251264  to the right, improve=0.08610527, (0 missing)
      WET   < 0.004161   to the left,  improve=0.01835295, (0 missing)
      AG    < 0.00991835 to the left,  improve=0.01495117, (0 missing)

Node number 2: 637 observations
  mean=0.1842229, MSE=0.08040125

Node number 3: 423 observations
  mean=0.5988999, MSE=0.6187896
```

ANALYSIS: TP VS. % DEVELOPED (nCPA)

cp	r2	mean left	mean right	pperm
0.1117865	0.1225523	0.1842229	0.5988999	0.001

5%	25%	50%	75%	95%
0.0812945	0.1117865	0.1173865	0.1622805	0.195147

ANALYSIS: TN VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TN ~ DEV + DEVAG + AG + FOR + WET, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=532 (1658 observations deleted due to missingness)
```

```
        CP nsplit rel error   xerror      xstd
1 0.04911466      0 1.0000000 1.003861 0.5595769
2 0.03821616      1 0.9508853 1.055182 0.5582847

Node number 1: 532 observations,    complexity param=0.04911466
  mean=2.594197, MSE=21.94937
  left son=2 (280 obs) right son=3 (252 obs)
  Primary splits:
      DEV   < 0.0797755 to the left,  improve=0.04911466, (0 missing)
      DEVAG < 0.465475  to the left,  improve=0.04375814, (0 missing)
      FOR   < 0.3056394 to the right, improve=0.02836732, (0 missing)
      AG    < 0.2244487 to the left,  improve=0.02062529, (0 missing)
      WET   < 0.0107185 to the left,  improve=0.02020472, (0 missing)

Node number 2: 280 observations
  mean=1.609194, MSE=3.216894

Node number 3: 252 observations
  mean=3.688646, MSE=40.48737
```

ANALYSIS: TN VS. %DEVELOPED (nCPA)

```
 cp              r2              mean left       mean right      pperm     5%      25%      50%
0.0797755       0.04911466      1.609194        3.688646        0.004 0.065229 0.07934 0.0819755
          75%      95%
[1,] 0.1391135 0.218493
```

ANALYSIS: CHLASPEC VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = CHLASPEC ~ DEV + AG + DEVAG + FOR + WET, data = streams,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=586 (1604 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.02347297      0 1.0000000 1.002740 0.1634325
2 0.02096011      4 0.9061081 1.011690 0.1631463

Node number 1: 586 observations,    complexity param=0.02347297
  mean=8.885287, MSE=51.11861
  left son=2 (568 obs) right son=3 (18 obs)
  Primary splits:
      WET   < 0.2595805 to the left,  improve=0.014537270, (0 missing)
      DEV   < 0.005749  to the left,  improve=0.012972250, (0 missing)
```

```
    FOR   < 0.7204692 to the right, improve=0.009985937, (0 missing)
    AG    < 0.000289  to the right, improve=0.005272042, (0 missing)
    DEVAG < 0.7227296 to the right, improve=0.003956970, (0 missing)

Node number 2: 568 observations,    complexity param=0.02347297
  mean=8.731828, MSE=48.71719
  left son=4 (46 obs) right son=5 (522 obs)
  Primary splits:
    DEV   < 0.005749  to the left,  improve=0.012574690, (0 missing)
    WET   < 0.0003795 to the left,  improve=0.009502900, (0 missing)
    FOR   < 0.7204692 to the right, improve=0.008673473, (0 missing)
    DEVAG < 0.2645541 to the left,  improve=0.006713071, (0 missing)
    AG    < 0.850807  to the left,  improve=0.004444450, (0 missing)

Node number 3: 18 observations
  mean=13.72778, MSE=102.704

Node number 4: 46 observations
  mean=6.095217, MSE=13.02579

Node number 5: 522 observations,    complexity param=0.02347297
  mean=8.964173, MSE=51.19582
  left son=10 (218 obs) right son=11 (304 obs)
  Primary splits:
    FOR   < 0.2495495 to the left,  improve=0.009944468, (0 missing)
    WET   < 0.167956  to the right, improve=0.008959459, (0 missing)
    DEV   < 0.0705025 to the right, improve=0.006856472, (0 missing)
    DEVAG < 0.7227296 to the right, improve=0.006007757, (0 missing)
    AG    < 0.850807  to the left,  improve=0.003758254, (0 missing)

Node number 10: 218 observations
  mean=8.121583, MSE=36.42679

Node number 11: 304 observations,    complexity param=0.02347297
  mean=9.568399, MSE=60.91256
  left son=22 (289 obs) right son=23 (15 obs)
  Primary splits:
    FOR   < 0.2962484 to the right, improve=0.095228750, (0 missing)
    DEVAG < 0.6335247 to the left,  improve=0.025054780, (0 missing)
    AG    < 0.6088519 to the left,  improve=0.022583180, (0 missing)
    DEV   < 0.3574215 to the left,  improve=0.019492340, (0 missing)
    WET   < 0.001393  to the left,  improve=0.009652881, (0 missing)

Node number 22: 289 observations
  mean=9.0197, MSE=43.08419
```

```
Node number 23: 15 observations
  mean=20.14, MSE=286.8464

ANALYSIS: CHLAFLUORO VS. WATERSHED LULC (CART)
Call:
mvpart(form = CHLAFLUORO ~ DEV + AG + DEVAG + FOR + WET, data = streams,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=345 (1845 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.03287493      0 1.0000000 1.007415 0.2087968
2 0.01798238      2 0.9342501 1.089604 0.2130477

Node number 1: 345 observations,    complexity param=0.03287493
  mean=10.17641, MSE=159.6029
  left son=2 (324 obs) right son=3 (21 obs)
  Primary splits:
      DEVAG < 0.9737215  to the left,  improve=0.03269615, (0 missing)
      WET   < 0.27771    to the left,  improve=0.02974471, (0 missing)
      AG    < 0.00044    to the right, improve=0.02378225, (0 missing)
      FOR   < 0.05512164 to the right, improve=0.02282121, (0 missing)
      DEV   < 0.992794   to the left,  improve=0.01850715, (0 missing)

Node number 2: 324 observations,    complexity param=0.03287493
  mean=9.594838, MSE=126.2962
  left son=4 (311 obs) right son=5 (13 obs)
  Primary splits:
      WET   < 0.27771    to the left,  improve=0.044477920, (0 missing)
      DEVAG < 0.4316418  to the right, improve=0.017263210, (0 missing)
      DEV   < 0.007653   to the left,  improve=0.014041220, (0 missing)
      AG    < 0.00044    to the right, improve=0.010179570, (0 missing)
      FOR   < 0.470028   to the left,  improve=0.009460488, (0 missing)

Node number 3: 21 observations
  mean=19.14929, MSE=587.7457

Node number 4: 311 observations
  mean=9.110265, MSE=99.79387

Node number 5: 13 observations
  mean=21.18731, MSE=620.3117
```

ANALYSIS: SECCHI VS. WATERSHED LULC

```
Call:
mvpart(form = SECCHI ~ DEV + AG + DEVAG + FOR + WET, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
```

No possible LULC splits for SECCHI

ANALYSIS: TP VS. UNWEIGHTED PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = TP ~ FLOW, data = streams, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=1064 (1126 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.16642355      0 1.0000000 1.0028394 0.1683075
2 0.01424009      1 0.8335764 0.8564614 0.1667769

Node number 1: 1064 observations,    complexity param=0.1664236
  mean=0.3490504, MSE=0.3353683
  left son=2 (704 obs) right son=3 (360 obs)
  Primary splits:
      FLOW < 3.6025 to the left,  improve=0.1664236, (0 missing)

Node number 2: 704 observations
  mean=0.1801101, MSE=0.157632

Node number 3: 360 observations
  mean=0.6794226, MSE=0.5179825Call:
mvpart(form = TP ~ FLOW, data = streams, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=1064 (1126 observations deleted due to missingness)
```

ANALYSIS: TP VS. UNWEIGHTED PERMITTED MUNICIPAL WWTP FLOW (nCPA)

cp	r2	mean left	mean right	pperm
3.6025	0.1672208	0.1800420	0.6815758	0.001

5%	25%	50%	75%	95%
3.016	3.4895	3.6025	5.11751	9.2625

ANALYSIS: TN VS. UNWEIGHTED PERMITTED MUNICIPAL WWTP FLOW (CART)

```
mvpart(form = TN ~ FLOW, data = str, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=535 (1655 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.04106422      0 1.0000000 1.0022664 0.5589595
2 0.02478708      1 0.9589358 0.9802297 0.5685007

Node number 1: 535 observations,    complexity param=0.04106422
  mean=2.584819, MSE=21.84203
  left son=2 (364 obs) right son=3 (171 obs)
  Primary splits:
      FLOW < 3.9725 to the left,  improve=0.04106422, (0 missing)

Node number 2: 364 observations
  mean=1.935698, MSE=23.10083

Node number 3: 171 observations
  mean=3.966573, MSE=16.35631
```

ANALYSIS: TN VS. UNWEIGHTED PERMITTED MUNICIPAL WWTP FLOW (nCPA)

```
         cp         r2 mean left mean right pperm      5%      25%    50%
[1,] 3.9725 0.04106422  1.935698   3.966573 0.005 0.4765 3.188125 3.9725
         75%      95%
[1,] 17.9125 132.2357
```

ANALYSIS: TP VS. AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = TP ~ WFLOW, data = streams, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=1061 (1129 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.32132569      0 1.0000000 1.0025343 0.1682600
2 0.01415295      1 0.6786743 0.7019645 0.1622207

Node number 1: 1061 observations,    complexity param=0.3213257
  mean=0.3492683, MSE=0.336293
  left son=2 (844 obs) right son=3 (217 obs)
  Primary splits:
      WFLOW < 0.03055805 to the left,  improve=0.3213257, (0 missing)

Node number 2: 844 observations
```

```
 mean=0.1825857, MSE=0.1346068

Node number 3: 217 observations
 mean=0.9975637, MSE=0.5923849
```

ANALYSIS: TP VS. AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW (nCPA)

cp	r2	mean left	mean right	pperm
0.03055805	0.3213257	0.1825857	0.9975637	0.001

5%	25%	50%	75%	95%
0.0278248	0.03055805	0.03413935	0.04471028	0.04849386

ANALYSIS: TN VS. AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = TN ~ WFLOW, data = streams, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=535 (1655 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.08517611      0 1.0000000 1.0046411 0.5601788
2 0.01000000      1 0.9148239 0.9367176 0.5677670

Node number 1: 535 observations,    complexity param=0.08517611
 mean=2.584819, MSE=21.84203
 left son=2 (445 obs) right son=3 (90 obs)
 Primary splits:
    WFLOW < 0.03055805 to the left,  improve=0.08517611, (0 missing)

Node number 2: 445 observations
 mean=1.971415, MSE=19.59679

Node number 3: 90 observations
 mean=5.617761, MSE=21.88436
```

ANALYSIS: TN VS. AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW (nCPA)

```
             cp         r2 mean left mean right pperm        5%        25%
[1,] 0.03055805 0.08517611  1.971415   5.617761 0.001 0.01076041 0.01288225

            50%        75%        95%
[1,] 0.03055805 0.03309036 0.06187353
```

ANALYSIS: CHLAPSEC VS. PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = CHLASPEC ~ FLOW + WFLOW, data = streams, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
  n=591 (1599 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.01599638      0 1.0000000 1.003934 0.1627978
2 0.01000000      2 0.9680072 1.051783 0.1634143

Node number 1: 591 observations,    complexity param=0.01599638
  mean=8.881266, MSE=50.82976
  left son=2 (83 obs) right son=3 (508 obs)
  Primary splits:
      WFLOW < 0.007219193 to the right, improve=0.01348964, (0 missing)
      FLOW  < 15.76805    to the right, improve=0.01253396, (0 missing)

Node number 2: 83 observations
  mean=6.832691, MSE=25.95088

Node number 3: 508 observations,    complexity param=0.01599638
  mean=9.215974, MSE=54.09691
  left son=6 (489 obs) right son=7 (19 obs)
  Primary splits:
      WFLOW < 0.004150783 to the left,  improve=0.020226200, (0 missing)
      FLOW  < 5.36        to the left,  improve=0.004301735, (0 missing)

Node number 6: 489 observations
  mean=9.009785, MSE=48.96514

Node number 7: 19 observations
  mean=14.52263, MSE=156.9177
```

ANALYSIS: CHLAFLUORO VS. PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ FLOW + WFLOW, data = streams, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
  n=349 (1841 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.01000553      0  1.000000 1.003430 0.2043948
2 0.01000000      2  0.979989 1.057189 0.2073541

Node number 1: 349 observations,    complexity param=0.01000553
```

```
  mean=10.31183, MSE=170.6774
  left son=2 (333 obs) right son=3 (16 obs)
  Primary splits:
      FLOW  < 24.415      to the left,  improve=0.006598594, (0 missing)
      WFLOW < 0.001087428 to the right, improve=0.005449407, (0 missing)

Node number 2: 333 observations,    complexity param=0.01000553
  mean=10.0792, MSE=167.355
  left son=4 (65 obs) right son=5 (268 obs)
  Primary splits:
      WFLOW < 0.001087428 to the right, improve=0.014335970, (0 missing)
      FLOW  < 0.10775     to the right, improve=0.007398835, (0 missing)

Node number 3: 16 observations
  mean=15.15328, MSE=215.2594

Node number 4: 65 observations
  mean=6.934038, MSE=63.60254

Node number 5: 268 observations
  mean=10.84202, MSE=189.5377
```

ANALYSIS: SECCHI VS. PERMITTED MUNICIPAL WWTP FLOW (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ FLOW + WFLOW, data = streams, xval = 10,
    method = "anova", minsplit = 20, minbucket = 10)

No splits possible
```

ANALYSIS: TP VS. REGIONS (CART)

```
Call:

mvpart(form = TP ~ BASIN + ECO3 + BASECO3, data = streams, xval = 10,
    method = "anova", minsplit = 20, minbucket = 10)
  n=1077 (1113 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.24480747      0 1.0000000 1.0015112 0.1676493
2 0.01585546      1 0.7551925 0.7961572 0.1390523

Node number 1: 1077 observations,    complexity param=0.2448075
  mean=0.3462392, MSE=0.3320902
  left son=2 (884 obs) right son=3 (193 obs)
```

Primary splits:
BASECO3 splits as LL-RLLLLRLLLRLLRLLLLLLLLLLLLLLRRLLLRLL-RLLLLRLLLLLLRL-LLLLL-LLLLR-LL, improve=0.2448075, (0 missing)
BASIN splits as LRLLLLLLLLLRLRLLLLLLLLL, improve=0.1420840, (0 missing)
ECO3 splits as LLRLLLLLLRRL, improve=0.1229040, (0 missing)

Node number 2: 884 observations
mean=0.213012, MSE=0.1260684

Node number 3: 193 observations
mean=0.9564611, MSE=0.8220649

ANALYSIS: TN VS. REGIONS (CART)

Call:
mvpart(form = TN ~ BASIN + ECO3 + BASECO3, data = streams, xval = 10,
method = "anova", minsplit = 20, minbucket = 10)
n=535 (1655 observations deleted due to missingness)

	CP	nsplit	rel error	xerror	xstd
1	0.13880121	0	1.0000000	1.001770	0.5586118
2	0.04125689	1	0.8611988	1.013992	0.5069495

Node number 1: 535 observations, complexity param=0.1388012
mean=2.584819, MSE=21.84203
left son=2 (469 obs) right son=3 (66 obs)
Primary splits:
BASECO3 splits as -L-RLLLLLLLLRLL-LLLLL--LLLLLLLRRLLL--L--LRL-RLLLL-LLL-LL-LL-LLLLR-LL, improve=0.13880120, (0 missing)
ECO3 splits as -LLLLLLLLRLL, improve=0.09238926, (0 missing)
BASIN splits as LRLRLLL-LRL-RRL-RLLLLRL, improve=0.06001182, (0 missing)

Node number 2: 469 observations
mean=1.931645, MSE=5.436439

Node number 3: 66 observations
mean=7.226311, MSE=113.846

ANALYSIS: CHLASPEC VS. REGIONS

Call:
mvpart(form = CHLASPEC ~ BASIN + ECO3 + BASECO3, data = streams,
xval = 10, method = "anova", minsplit = 20, minbucket = 10)
n=597 (1593 observations deleted due to missingness)

```
        CP nsplit rel error   xerror      xstd
1 0.06813607      0  1.000000 1.005460 0.1624268
2 0.02852109      1  0.931864 1.077224 0.1645635

Node number 1: 597 observations,    complexity param=0.06813607
  mean=8.80926, MSE=50.8417
  left son=2 (427 obs) right son=3 (170 obs)
  Primary splits:
      BASECO3 splits as  LLLRLL-LLLLLLLLLLLLLLRLLRRRLRRRLRLLLRLLRRLLRRLRLRRLRLRRLRLL-RRLLRRL-, improve=0.06813607, (0 missing)
      BASIN   splits as  LLLLLLLLLLLLLLLRLRLLLL-, improve=0.04415343, (0 missing)
      ECO3    splits as  LRLRLLLLLRRR, improve=0.01794812, (0 missing)

Node number 2: 427 observations
  mean=7.634879, MSE=35.64969

Node number 3: 170 observations
  mean=11.75903, MSE=76.83514
```

ANALYSIS: CHLAFLUORO VS. REGIONS

```
Call:
mvpart(form = CHLAFLUORO ~ BASIN + ECO3 + BASECO3, data = streams,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=352 (1838 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.15456303      0  1.000000 1.004399 0.2045104
2 0.04141581      1  0.845437 1.036678 0.2080253

Node number 1: 352 observations,    complexity param=0.154563
  mean=10.29809, MSE=169.2806
  left son=2 (319 obs) right son=3 (33 obs)
  Primary splits:
      BASECO3 splits as  LL-LLL-LLLLLLLL--LLLLLRLLL-L-LL-RL---LLRLL-RLLLLRRLLLLR-LLL-LLLLLLLL, improve=0.15456300, (0 missing)
      BASIN   splits as  LLLLLLLLLLRRLLRRLRLRLLL, improve=0.06751251, (0 missing)
      ECO3    splits as  LR-RLLLLRLRR, improve=0.05497201, (0 missing)

Node number 2: 319 observations
  mean=8.652892, MSE=112.3894

Node number 3: 33 observations
  mean=26.20167, MSE=440.1413
```

ANALYSIS: SECCHI TRANSPARENCY VS. REGIONS

```
Call:
mvpart(form = SECCHI ~ BASIN + ECO3 + BASECO3, data = streams,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=977 (1213 observations deleted due to missingness)

          CP nsplit rel error   xerror       xstd
1 0.04331268      0 1.0000000 1.002235 0.06046411
2 0.01200963      1 0.9566873 1.021130 0.06128966

Node number 1: 977 observations,    complexity param=0.04331268
  mean=0.519675, MSE=0.1113098
  left son=2 (577 obs) right son=3 (400 obs)
  Primary splits:
      BASECO3 splits as  LR-RRRRLRLLLRLR-LLRRLLLRLRLLLLLLLRLLRRRLLRRLLRLLLRRLLLLLLLRLLLLLRRRL, 
improve=0.043312680, (0 missing)
      BASIN   splits as  RRRRRRLLRLRRRLLRLLLLLLL, improve=0.011842170, (0 missing)
      ECO3    splits as  LRRRRRRRLRRR, improve=0.006165701, (0 missing)

Node number 2: 577 observations
  mean=0.4618633, MSE=0.08691248

Node number 3: 400 observations
  mean=0.6030685, MSE=0.1347272
```

Appendix 1.11. Classification and regression tree (CART) and non-parametric changepoint analysis code from stressor-response analyses on the water quality median database

ANALYSIS: SECCHI TRANSPARENCY VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "ALL" STATIONS DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TN + TP + NO3 + NH4 + SRP, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=911 (1279 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.30827367      0 1.0000000 1.0011323 0.06216640
2 0.07171672      1 0.6917263 0.8638002 0.04616456

Node number 1: 911 observations,    complexity param=0.3082737
  mean=0.5200763, MSE=0.1147619
  left son=2 (565 obs) right son=3 (253 obs), 93 observations remain
  Primary splits:
      TP  < 0.06325 to the right, improve=0.20417800, (93 missing)
      SRP < 0.04725 to the right, improve=0.06512497, (77 missing)
      NH4 < 0.05075 to the right, improve=0.05286194, (156 missing)
      TN  < 0.69625 to the right, improve=0.02301554, (406 missing)
      NO3 < 0.4525  to the left,  improve=0.01035787, (143 missing)

Node number 2: 565 observations
  mean=0.4145898, MSE=0.05612627

Node number 3: 253 observations
  mean=0.7640958, MSE=0.1605033
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE "ALL" STATIONS DATASET (nCPA)

```
          cp        r2 mean left mean right pperm      5%     25%   50%    75%
[1,] 0.06325 0.2279015 0.7640958  0.4145898 0.001 0.06175 0.06325 0.064 0.0685
        95%
[1,] 0.0848
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE "ALL" STATIONS DATASET (nCPA)

```
          cp         r2 mean left mean right pperm      5%     25%  50%   75%
[1,] 0.69625 0.03747921 0.7258203  0.5183265 0.003 0.56335 0.69625 0.76 0.895
          95%
[1,] 3.501862
```

ANALYSIS: SECCHI TRANSPARENCY VS. NOX-N IN THE "ALL" STATIONS DATASET (nCPA)

```
      cp        r2 mean left mean right pperm     5%    25%   50%  75%   95%
[1,] 0.4525 0.01161302 0.5024879  0.5791453 0.068 0.0332 0.4525 0.835 3.51 6.742
```

ANALYSIS: SECCHI TRANSPARENCY VS. NH4-N IN THE "ALL" STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%     25%     50%   75%     95%
[1,] 0.05075 0.06440622 0.6010463  0.4298225 0.001 0.05075 0.05075 0.05075 0.051 0.05425
```

ANALYSIS: SECCHI TRANSPARENCY VS. SRP IN THE "ALL" STATIONS DATASET (nCPA)

```
       cp       r2 mean left mean right pperm     5%     25%     50%    75%    95%
[1,] 0.04725 0.0729688    0.6226   0.439136 0.001 0.0425 0.04725 0.04725 0.0505 0.2435
```

ANALYSIS: SECCHI TRANSPARENCY VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "LOW" AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TN + TP + NO3 + NH4 + SRP, data = low,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=715 (1076 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.30461028      0 1.0000000 1.0013663 0.06653440
2 0.07821683      1 0.6953897 0.8924197 0.05137561

Node number 1: 715 observations,    complexity param=0.3046103
  mean=0.5471792, MSE=0.1243952
  left son=2 (382 obs) right son=3 (246 obs), 87 observations remain
  Primary splits:
      TP  < 0.06325 to the right, improve=0.18951450, (87 missing)
      SRP < 0.04725 to the right, improve=0.04015748, (67 missing)
      NH4 < 0.05075 to the right, improve=0.03106828, (142 missing)
      NO3 < 0.4525  to the left,  improve=0.02484058, (64 missing)
      TN  < 0.69625 to the right, improve=0.02287423, (297 missing)

Node number 2: 382 observations
  mean=0.4245205, MSE=0.05843551

Node number 3: 246 observations
  mean=0.7601474, MSE=0.1606804
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
        cp       r2 mean left mean right pperm     5%    25%     50%    75%
[1,] 0.06325 0.2141639 0.7601474  0.4245205 0.001 0.06175 0.0625 0.06325 0.0685
       95%
[1,] 0.0848
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm      5%      25%    50%      75%    95%
[1,] 0.69625 0.03666493 0.7258203   0.532078 0.011 0.552575 0.636375 0.7268 0.841375 3.9565
```

ANALYSIS: SECCHI TRANSPARENCY VS. NO_X-N IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
      cp        r2 mean left mean right pperm      5%    25%  50%    75%      95%
[1,] 0.4525 0.02679840 0.5076687  0.6322914 0.005 0.257375 0.4525 0.56 0.6975 1.827875
```

ANALYSIS: SECCHI TRANSPARENCY VS. NH_4-N IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%     25%     50%  75%    95%
[1,] 0.05075 0.03846916  0.606153  0.4624554 0.001 0.0205 0.05075 0.05075 0.051 0.0525
```

ANALYSIS: SECCHI TRANSPARENCY VS. SRP IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
        cp       r2 mean left mean right pperm     5%     25%     50%    75%    95%
[1,] 0.04725 0.0455758 0.6152305  0.4666755 0.001 0.0425 0.04725 0.04725 0.0475 0.0525
```

ANALYSIS: SECCHI TRANSPARENCY VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "HIGH" AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TN + TP + NO3 + NH4 + SRP, data = high,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=196 (203 observations deleted due to missingness)

        CP nsplit rel error   xerror     xstd
1 0.20468998      0   1.00000 1.0114221 0.1432405
2 0.08926737      1   0.79531 0.9124827 0.1229194

Node number 1: 196 observations,    complexity param=0.20469
  mean=0.4212061, MSE=0.06716516
  left son=2 (175 obs) right son=3 (11 obs), 10 observations remain
  Primary splits:
     SRP < 0.045   to the right, improve=0.16393370, (10 missing)
     TP  < 0.126    to the right, improve=0.16161590, (6 missing)
     NH4 < 0.0605  to the right, improve=0.05513448, (14 missing)
```

```
    TN  < 2.43675 to the right, improve=0.04979814, (109 missing)
    NO3 < 1.075   to the left,  improve=0.01826452, (79 missing)

Node number 2: 175 observations
  mean=0.3897223, MSE=0.05387172

Node number 3: 11 observations
  mean=0.8463636, MSE=0.09474587
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%   25%    50%      75%
[1,] 0.091 0.1794622    0.8575  0.3850943 0.001 0.085 0.096 0.1285 0.337125
        95%
[1,] 0.5725
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
          cp         r2 mean left mean right pperm    5%    25%     50%     75%   95%
[1,] 2.43675 0.08192565 0.6071739  0.4103344  0.12 2.195 2.3375 2.43675 3.04175 7.435
```

ANALYSIS: SECCHI TRANSPARENCY VS. NOX-N IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
        cp         r2 mean left mean right pperm    5%   25%     50%   75%     95%
[1,] 1.075 0.02434734 0.3802941  0.4801373 0.598 0.818 1.075 2.12025 3.805 8.07025
```

ANALYSIS: SECCHI TRANSPARENCY VS. NH4-N IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
         cp         r2 mean left mean right pperm      5%    25%     50%    75%   95%
[1,] 0.0605 0.06485957 0.5401143  0.3798803 0.019 0.05175 0.0605 0.07425 0.1075 0.108
```

ANALYSIS: SECCHI TRANSPARENCY VS. SRP IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%  25%    50%   75%
[1,] 0.05 0.1778172 0.8209091    0.38992 0.001 0.045 0.05 0.0635 0.073
          95%
[1,] 0.1911375
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE “ALL” STATIONS DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TN + TP + NO3 + NH4 + SRP, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=597 (1593 observations deleted due to missingness)
```

```
         CP nsplit rel error    xerror      xstd
1 0.1332654      0 1.0000000 1.0036420 0.1615709
2 0.1025975      1 0.8667346 0.9565567 0.1498199

Node number 1: 597 observations,    complexity param=0.1332654
  mean=8.80926, MSE=50.8417 
  left son=2 (298 obs) right son=3 (297 obs), 2 observations remain
  Primary splits:
      TP  < 0.09916668 to the left,  improve=0.08525244, (2 missing)
      SRP < 0.03525    to the left,  improve=0.03799185, (56 missing)
      NH4 < 0.0495     to the left,  improve=0.03217101, (41 missing)
      TN  < 1.1075     to the left,  improve=0.02579898, (198 missing)
      NO3 < 0.0225     to the right, improve=0.01368464, (63 missing)

Node number 2: 298 observations
  mean=6.638809, MSE=19.62364 

Node number 3: 297 observations,    complexity param=0.1025975
  mean=10.80964, MSE=68.88791 
  left son=6 (251 obs) right son=7 (21 obs), 25 observations remain
  Primary splits:
      NO3 < 0.0633     to the right, improve=0.055900600, (25 missing)
      SRP < 0.0855     to the right, improve=0.032941390, (15 missing)
      TN  < 7.0545     to the right, improve=0.023445540, (99 missing)
      NH4 < 0.0915     to the right, improve=0.022980860, (14 missing)
      TP  < 0.1108332  to the left,  improve=0.005086337, (0 missing)

Node number 6: 251 observations
  mean=10.04976, MSE=47.59053 

Node number 7: 21 observations
  mean=17.73214, MSE=257.1618 
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE "ALL" STATIONS DATASET (nCPA)

```
             cp        r2 mean left mean right pperm     5%   25%   50%
[1,] 0.09916668 0.0895521  6.638809   10.80964 0.001 0.0655 0.091 0.098
           75%    95%
[1,] 0.1108332 0.1125
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE "ALL" STATIONS DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%    25%    50%    75%      95%
[1,] 1.1075 0.04287716  7.938434   10.78062  0.01 0.745 1.0559 1.1075 1.1675 1.945737
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NO_X-N IN THE "ALL" STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%    25%     50%      75%   95%
[1,] 0.0225 0.01548387  14.34321   8.823471 0.138 0.0225 0.0225 0.04635 0.651875 7.725
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NH_4-N IN THE "ALL" STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm      5%   25%    50%    75%   95%
[1,] 0.0495 0.03558166  6.417033   9.609823 0.001 0.04575 0.049 0.0495 0.0495 0.094
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. SRP IN THE "ALL" STATIONS DATASET (nCPA)

```
        cp       r2 mean left mean right pperm      5%     25%   50%    75%    95%
[1,] 0.03525 0.0442112  5.163898   9.847624 0.005 0.03275 0.03525 0.037 0.0425 0.0585
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "LOW" AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP, data = low, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=497 (1294 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.10648028      0 1.0000000 1.0034242 0.1869510
2 0.01793165      1 0.8935197 0.9024399 0.1639363

Node number 1: 497 observations,    complexity param=0.1064803
  mean=8.699393, MSE=51.2273
  left son=2 (312 obs) right son=3 (185 obs)
  Primary splits:
      TP < 0.1108332 to the left,  improve=0.1064803, (0 missing)

Node number 2: 312 observations
  mean=6.900962, MSE=20.00556

Node number 3: 185 observations
  mean=11.73242, MSE=89.22837
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%    25%    50%
[1,] 0.1108332 0.08593495  7.169855   11.46970 0.001 0.0655 0.0685 0.1085
          75%   95%
[1,] 0.1108332 0.114
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
      cp        r2 mean left mean right pperm      5%    25%   50%    75%  95%
[1,] 1.1075 0.06005272  7.945957   11.36896 0.005 0.794875 1.0925 1.148 1.2945 2.02
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NO_X-N IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
      cp        r2 mean left mean right pperm     5%    25%    50%    75%   95%
[1,] 0.0225 0.01800935  14.34321   8.730012 0.128 0.0225 0.0225 0.0377 0.8975 5.954
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NH_4-N IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm      5%   25%     50%    75%     95%
[1,] 0.04875 0.02892779  6.742207   9.615946 0.016 0.022005 0.048 0.04875 0.0915 0.09475
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. SRP IN THE "LOW" WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm      5%     25%    50%    75%      95%
[1,] 0.03525 0.04891814  5.218448    9.95717 0.006 0.03275 0.03525 0.0375 0.0505 0.058825
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "HIGH" AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TN + TP + NO3 + NH4 + SRP, data = high,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=98 (301 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.12469662      0 1.0000000 1.021787 0.2242717
2 0.07574669      2 0.7506068 1.144527 0.2175474

Node number 1: 98 observations,    complexity param=0.1246966
  mean=8.828878, MSE=35.03915
  left son=2 (65 obs) right son=3 (33 obs)
  Primary splits:
      NH4 < 0.1075  to the left,  improve=0.11822050, (0 missing)
      TP  < 0.15825 to the left,  improve=0.09909359, (0 missing)
      NO3 < 1.165   to the right, improve=0.08490351, (7 missing)
      TN  < 3.175   to the left,  improve=0.03559067, (23 missing)
      SRP < 0.435   to the left,  improve=0.02731838, (9 missing)

Node number 2: 65 observations
  mean=7.378692, MSE=23.53012
```

```
Node number 3: 33 observations,    complexity param=0.1246966
  mean=11.6853, MSE=45.40694
  left son=6 (16 obs) right son=7 (15 obs), 2 observations remain
  Primary splits:
      SRP < 0.655   to the right, improve=0.15734590, (2 missing)
      NH4 < 0.24    to the right, improve=0.12823670, (0 missing)
      TP  < 1       to the right, improve=0.10560240, (0 missing)
      TN  < 5.14    to the right, improve=0.09492327, (8 missing)
      NO3 < 3.865   to the right, improve=0.07140823, (1 missing)

Node number 6: 16 observations
  mean=9.6625, MSE=3.883594

Node number 7: 15 observations
  mean=15.181, MSE=65.72439
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
          cp         r2 mean left mean right pperm      5%       25%   50%       75%       95%
[1,] 0.66425 0.04537682   7.95375   10.44615 0.306 0.14425 0.1854375 0.395 0.5555626 0.718762
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
       cp         r2 mean left mean right pperm      5%     25%     50%    75%      95%
[1,] 3.21 0.03718881  7.937759   10.17534 0.562 1.59775 2.80075 3.2825 4.7205 8.26875
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NO_X-N IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%   25%   50%  75%   95%
[1,] 1.165 0.09415435  12.34174   8.223088 0.056 0.8225 0.909 1.165 1.23 3.985
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. NH_4-N IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
         cp        r2 mean left mean right pperm    5%    25%  50%     75%    95%
[1,] 0.0495 0.1500157  3.409167   10.13087 0.005 0.045 0.0495 0.05 0.1075 0.145
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. SRP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
          cp         r2 mean left mean right pperm      5%       25%   50%       75%       95%
[1,] 0.66425 0.04537682   7.95375   10.44615 0.306 0.14425 0.1854375 0.395 0.5555626 0.718762
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "ALL" STATIONS DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TN + TP + NO3 + NH4 + SRP, data = streams,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=352 (1838 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.08922706      0 1.0000000 1.002487 0.2036106
2 0.04819618      5 0.5466419 1.251350 0.1885144

Node number 1: 352 observations,    complexity param=0.08922706
  mean=10.29809, MSE=169.2806
  left son=2 (127 obs) right son=3 (212 obs), 13 observations remain
  Primary splits:
      TP  < 0.0685    to the left,  improve=0.06625868, (13 missing)
      TN  < 1.1325    to the left,  improve=0.06149559, (55 missing)
      NH4 < 0.04875   to the left,  improve=0.02865018, (46 missing)
      SRP < 0.0505    to the left,  improve=0.01841437, (7 missing)
      NO3 < 0.345     to the right, improve=0.01521774, (10 missing)

Node number 2: 127 observations
  mean=5.982402, MSE=30.01164

Node number 3: 212 observations,    complexity param=0.08922706
  mean=13.03301, MSE=242.2402
  left son=6 (37 obs) right son=7 (147 obs), 28 observations remain
  Primary splits:
      TN  < 1.1325    to the left,  improve=0.04428050, (28 missing)
      SRP < 0.0702    to the right, improve=0.04380472, (5 missing)
      NH4 < 0.04875   to the left,  improve=0.03869686, (17 missing)
      NO3 < 0.3452    to the right, improve=0.03823956, (6 missing)
      TP  < 1.1525    to the right, improve=0.01146242, (0 missing)

Node number 6: 37 observations
  mean=6.457297, MSE=23.88775

Node number 7: 147 observations,    complexity param=0.08922706
  mean=15.22825, MSE=312.9351
  left son=14 (93 obs) right son=15 (51 obs), 3 observations remain
  Primary splits:
      SRP < 0.0702    to the right, improve=0.14409480, (3 missing)
      NO3 < 0.2320833 to the right, improve=0.14125900, (0 missing)
      TP  < 0.185     to the right, improve=0.06410112, (0 missing)
```

```
    NH4 < 0.0475    to the left,  improve=0.05681256, (13 missing)
    TN  < 7         to the right, improve=0.03931012, (0 missing)

Node number 14: 93 observations
  mean=10.4511, MSE=120.526

Node number 15: 51 observations,    complexity param=0.08922706
  mean=24.63725, MSE=543.7702
  left son=30 (28 obs) right son=31 (23 obs)
  Primary splits:
    TP  < 0.1225    to the left,  improve=0.115637200, (0 missing)
    TN  < 1.905     to the left,  improve=0.103681700, (0 missing)
    NO3 < 0.3175    to the right, improve=0.095814380, (0 missing)
    SRP < 0.045     to the right, improve=0.006969917, (0 missing)
    NH4 < 0.065     to the left,  improve=0.004540766, (3 missing)

Node number 30: 28 observations
  mean=17.45036, MSE=361.184

Node number 31: 23 observations,    complexity param=0.08922706
  mean=33.38652, MSE=626.6192
  left son=62 (13 obs) right son=63 (10 obs)
  Primary splits:
    TN  < 1.5025    to the left,  improve=0.545161700, (0 missing)
    TP  < 0.1815    to the right, improve=0.046310090, (0 missing)
    NO3 < 0.2225    to the left,  improve=0.045175860, (0 missing)
    NH4 < 0.0525    to the right, improve=0.003746705, (0 missing)

Node number 62: 13 observations
  mean=17.17615, MSE=109.0569

Node number 63: 10 observations
  mean=54.46, MSE=513.7499
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE “ALL” STATIONS DATASET (nCPA)

```
[1,] 0.0685 0.06678798  5.982402   13.03301 0.001 0.0675 0.0685 0.07875
        75%    95%
[1,] 0.091725 0.1225
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE “ALL” STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm   5%    25%   50%   75%
[1,] 1.1325 0.04625819  6.457297   15.22825 0.059 1.06 1.1325 1.145 1.245
      95%
```

```
[1,] 6.795
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NOX-N IN THE "ALL" STATIONS DATASET (nCPA)

```
      cp        r2 mean left mean right pperm    5%    25%   50%    75%    95%
[1,] 0.345 0.01539166  11.97031   8.713634 0.277 0.055 0.2115 0.345 1.8525 5.7225
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NH4-N IN THE "ALL" STATIONS DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%     25%    50%     75%    95%
[1,] 0.04875 0.02954198  5.678627   12.01651  0.03 0.0455 0.04875 0.0525 0.145625 0.2075
```

ANALYSIS: FLUOROMETRIC CHL-A VS. SRP IN THE "ALL" STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%     25%    50%    75%   95%
[1,] 0.0505 0.01851426  8.472038   12.05328 0.125 0.0475 0.04975 0.0505 0.0575 0.655
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE "LOW" AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TN + TP + NO3 + NH4 + SRP, data = low,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=292 (1499 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.09285062      0 1.0000000 1.006042 0.2316881
2 0.07055112      3 0.7214482 1.352173 0.2419422

Node number 1: 292 observations,    complexity param=0.09285062
  mean=10.22914, MSE=177.012
  left son=2 (134 obs) right son=3 (146 obs), 12 observations remain
  Primary splits:
      TP  < 0.07875   to the left,  improve=0.08139614, (12 missing)
      TN  < 1.14575   to the left,  improve=0.07654235, (49 missing)
      NH4 < 0.04875   to the left,  improve=0.02611135, (39 missing)
      SRP < 0.0505    to the left,  improve=0.02064438, (5 missing)
      NO3 < 0.345     to the right, improve=0.01909529, (7 missing)

Node number 2: 134 observations
  mean=6.280112, MSE=36.70089

Node number 3: 146 observations,    complexity param=0.09285062
  mean=14.03981, MSE=288.4107
  left son=6 (31 obs) right son=7 (92 obs), 23 observations remain
```

```
  Primary splits:
      TN  < 1.1325    to the left,  improve=0.05980741, (23 missing)
      NO3 < 0.0625    to the right, improve=0.04815061, (3 missing)
      SRP < 0.0702    to the right, improve=0.04295914, (3 missing)
      NH4 < 0.04875   to the left,  improve=0.04122547, (11 missing)
      TP  < 0.1865    to the right, improve=0.01458294, (0 missing)

Node number 6: 31 observations
  mean=6.846452, MSE=25.7037

Node number 7: 92 observations,    complexity param=0.09285062
  mean=17.26813, MSE=398.4644
  left son=14 (66 obs) right son=15 (26 obs)
  Primary splits:
      NO3 < 0.2320833 to the right, improve=0.138673600, (0 missing)
      SRP < 0.0702    to the right, improve=0.129170000, (2 missing)
      TP  < 0.1865    to the right, improve=0.066780950, (0 missing)
      NH4 < 0.0475    to the left,  improve=0.058492800, (8 missing)
      TN  < 1.9475    to the left,  improve=0.009717242, (0 missing)

Node number 14: 66 observations
  mean=12.60254, MSE=256.5789

Node number 15: 26 observations
  mean=29.11154, MSE=563.1122
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
      cp         r2 mean left mean right pperm     5%   25%   50%      75%
[1,] 0.07875 0.08211816  6.280112   14.03981 0.001 0.0675 0.075 0.085 0.091725
       95%
[1,] 0.1225
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%   25%  50%    75%
[1,] 1.1325 0.04012506  8.099552   14.59102 0.084 1.125 1.145 1.24 1.4825
      95%
[1,] 3.915
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NO$_X$-N IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
      cp         r2 mean left mean right pperm     5%   25%    50%    75%  95%
[1,] 0.345 0.01933917  11.85032    8.08533 0.208 0.0525 0.165 0.3375 0.3452 2.14
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NH_4-N IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%     25%     50%    75%    95%
[1,] 0.04875 0.02694555  5.809535   11.95881 0.051 0.0455 0.04875 0.04875 0.0525 0.0825
```

ANALYSIS: FLUOROMETRIC CHL-A VS. SRP IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%     25%   50%     75%      95%
[1,] 0.0505 0.02075571  8.553007   12.41789  0.13 0.0475 0.04975 0.055 0.263125 0.440625
```

ANALYSIS: FLUOROMETRIC CHL-A VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) IN THE “HIGH” AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TN + TP + NO3 + NH4 + SRP, data = high,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=60 (339 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.22143540      0 1.0000000 1.046848 0.3221202
2 0.03273015      1 0.7785646 1.095931 0.2635095

Node number 1: 60 observations,    complexity param=0.2214354
  mean=10.63367, MSE=131.5187
  left son=2 (37 obs) right son=3 (16 obs), 7 observations remain
  Primary splits:
      NH4 < 0.145  to the left,  improve=0.19689860, (7 missing)
      SRP < 0.7675 to the right, improve=0.07660632, (2 missing)
      NO3 < 1.17   to the right, improve=0.06736672, (3 missing)
      TP  < 1.035  to the right, improve=0.05257248, (1 missing)
      TN  < 5.1475 to the right, improve=0.04677517, (6 missing)

Node number 2: 37 observations
  mean=7.621757, MSE=69.7655

Node number 3: 16 observations
  mean=19.41594, MSE=222.6516
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE “HIGH” WWTP FLOW DATASET (nCPA)

```
      cp        r2 mean left mean right pperm    5%    25%  50%  75%    95%
[1,] 0.71 0.05363386  7.574583   12.40259 0.494 0.195 0.5395 0.71 1.03 1.0925
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%   25%    50%    75%     95%
[1,] 5.1475 0.04721382   13.7352   8.491897 0.478 2.4025 3.505 5.1475 6.1915 7.091437
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
      cp        r2 mean left mean right pperm      5%  25%  50%     75%   95%
[1,] 1.17 0.06757603  15.62406   8.827683 0.342 0.737175 1.17 1.41 4.461875 5.865
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
       cp      r2 mean left mean right pperm    5%   25%   50%   75%    95%
[1,] 0.145 0.200639  7.526618   19.41594 0.011 0.105 0.108 0.145 0.145 0.1975
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE "HIGH" WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%    25%     50% 75%  95%
[1,] 0.7675 0.07703128  13.36743   6.767826 0.204 0.1075 0.4375 0.71925 0.8 1.01
```

ANALYSIS: 24-HOUR DO FLUX VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP) AND CHL-A IN THE "ALL" STATIONS DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TN + TP + NO3 + NH4 + SRP + CHLASPEC +
    CHLAFLUORO, data = streams, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=82 (2108 observations deleted due to missingness)

        CP nsplit rel error   xerror     xstd
1 0.540938      0  1.000000 1.034378 0.3482261
2 0.010000      1  0.459062 1.892490 0.3549068

Node number 1: 82 observations,    complexity param=0.540938
  mean=1.944329, MSE=2.722211
  left son=2 (21 obs) right son=3 (13 obs), 48 observations remain
  Primary splits:
      CHLAFLUORO < 5.375   to the left,  improve=0.26273840, (48 missing)
      NO3        < 0.0625  to the right, improve=0.09179852, (17 missing)
      TN         < 0.97    to the left,  improve=0.08865163, (38 missing)
      TP         < 0.06025 to the right, improve=0.05883695, (3 missing)
      CHLASPEC   < 7.6525  to the left,  improve=0.02133007, (37 missing)

Node number 2: 21 observations
  mean=1.179286, MSE=0.5954269
```

```
Node number 3: 13 observations
  mean=3.881923, MSE=6.920652
```

ANALYSIS: 24-HOUR DO FLUX VS. FLUOROMETRIC CHL-A IN THE “ALL” STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%  25%    50%   75%
[1,] 5.375 0.3640042  1.179286   3.881923 0.002 3.7475 4.55 5.3475 5.375
       95%
[1,] 9.9585
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE “ALL” STATIONS DATASET (nCPA)

```
         cp         r2 mean left mean right pperm   5%     25%     50%  75%    95%
[1,] 0.06025 0.05907337    2.6305    1.692797 0.214 0.06 0.06025 0.06175 0.09 0.21225
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE “ALL” STATIONS DATASET (nCPA)

```
      cp        r2 mean left mean right pperm     5%  25%    50%    75%    95%
[1,] 0.97 0.1144902     1.224   2.638793 0.133 0.9075 0.97 0.9775 1.0325 1.2175
```

ANALYSIS: 24-HOUR DO FLUX VS. NO_X-N IN THE “ALL” STATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm      5%    25%   50%  75%  95%
[1,] 0.055 0.1441459  3.031538   1.820865 0.038 0.04135 0.0427 0.055 0.08 0.25
```

ANALYSIS: 24-HOUR DO FLUX VS. NH_4-N IN THE “ALL” STATIONS DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%    25%    50%    75%   95%
[1,] 0.0575 0.02208717  2.305606   1.780455 0.682 0.0335 0.0525 0.0605 0.0795 0.103
```

ANALYSIS: 24-HOUR DO FLUX VS. SRP IN THE “ALL” STATIONS DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%    25%     50%   75%     95%
[1,] 0.0375 0.02011717  1.489167   2.136429 0.796 0.0375 0.0375 0.04725 0.065 0.18275
```

ANALYSIS: 24-HOUR DO FLUX VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP)AND CHL-A IN THE “LOW” AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TN + TP + NO3 + NH4 + SRP + CHLASPEC +
   CHLAFLUORO, data = low, xval = 10, method = "anova", minsplit = 20,
   minbucket = 10)
  n=72 (1719 observations deleted due to missingness)
```

```
        CP nsplit rel error   xerror      xstd
1 0.5452612      0 1.0000000 1.032989 0.3926650
2 0.0100000      1 0.4547388 1.888308 0.2980144

Node number 1: 72 observations,    complexity param=0.5452612
  mean=1.833472, MSE=2.809957
  left son=2 (19 obs) right son=3 (10 obs), 43 observations remain
  Primary splits:
      CHLAFLUORO < 5.375   to the left,  improve=0.28704520, (43 missing)
      NO3        < 0.0625  to the right, improve=0.12411960, (15 missing)
      TN         < 0.97    to the left,  improve=0.09462573, (35 missing)
      TP         < 0.06025 to the right, improve=0.08760556, (2 missing)
      NH4        < 0.0575  to the right, improve=0.05958105, (15 missing)

Node number 2: 19 observations
  mean=1.110263, MSE=0.5080144

Node number 3: 10 observations
  mean=4.0875, MSE=8.234906
```

ANALYSIS: 24-HOUR DO FLUX VS. FLUOROMETRIC CHL-A IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
     cp        r2 mean left mean right pperm    5%  25% 50%   75%    95%
[1,]  5 0.2984157    1.0625      3.895 0.001 3.395 4.47   5 5.055 5.5715
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
          cp       r2 mean left mean right pperm   5%     25%     50%     75%    95%
[1,] 0.06025 0.087747  2.650263   1.518725 0.097 0.06 0.06025 0.06025 0.07625 0.1675
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%  25%    50%    75%    95%
[1,] 0.97 0.1185712     1.224   2.689091 0.139 0.9075 0.97 0.9775 1.0325 1.1205
```

ANALYSIS: 24-HOUR DO FLUX VS. NO_X-N IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
        cp        r2 mean left mean right pperm      5%   25%   50%  75%    95%
[1,] 0.055 0.1788189  3.031538   1.658182 0.021 0.04135 0.045 0.055 0.08 0.1205
```

ANALYSIS: 24-HOUR DO FLUX VS. NH_4-N IN THE “LOW” WWTP FLOW DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%   25%    50%    75%    95%
[1,] 0.0575 0.06436263  2.353871   1.430577 0.186 0.045 0.055 0.0575 0.0795 0.0795
```

ANALYSIS: 24-HOUR DO FLUX VS. SRP IN THE “LOW” WWTP FLOW DATASET (nCPA)

	cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
[1,]	0.04725	0.04406974	2.306607	1.558387	0.371	0.0375	0.04475	0.04725	0.0525	0.0775

ANALYSIS: 24-HOUR DO FLUX VS. NUTRIENT STRESSORS (TP, TN, NOX-N, NH4-N, SRP)AND CHL-A IN THE “HIGH” AREA-WEIGHTED PERMITTED MUNICIPAL WWTP FLOW DATASET (CART)

No splits possible

Appendix 1.12. Classification and regression tree (CART) and non-parametric changepoint analysis code from stressor-response analyses on the bioassessment database

ANALYSIS: FISH IBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = FISHIBI ~ TP + TN + HQI, data = bioall,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=170 (352 observations deleted due to missingness)

          CP nsplit rel error   xerror       xstd
1 0.58349871      0 1.0000000 1.008847 0.09135959
2 0.07162249      1 0.4165013 1.123314 0.10591238

Node number 1: 170 observations,    complexity param=0.5834987
  mean=42.12059, MSE=45.74575
  left son=2 (48 obs) right son=3 (52 obs), 70 observations remain
  Primary splits:
      TN  < 1.4125  to the right, improve=0.11588370, (70 missing)
      TP  < 0.08425 to the right, improve=0.09245590, (46 missing)
      HQI < 21.125  to the left,  improve=0.06421473, (3 missing)

Node number 2: 48 observations
  mean=39.13542, MSE=31.8202

Node number 3: 52 observations
  mean=45.14423, MSE=32.9167
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (nCPA)

```
         cp        r2 mean left mean right pperm  5%    25%     50%    75%    95%
[1,] 21.125 0.0841236  40.80909   44.92727 0.003 15.5 18.125 19.4375 21.125 23.375
```

ANALYSIS: FISH IBI VS.TP IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (nCPA)

```
          cp        r2 mean left mean right pperm     5%    25%     50%    75%     95%
[1,] 0.08425 0.1339880  45.14706   40.25342 0.002 0.0725 0.0795 0.08425 0.1125 0.59425
```

ANALYSIS: FISH IBI VS. TN IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (nCPA)

```
         cp       r2 mean left mean right pperm  5%    25%   50%     75%   95%
[1,] 1.4125 0.217669  45.14423   39.13542 0.001 1.17 1.3975 1.465 1.90775 2.955
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = FISHIBI ~ TP + TN + HQI, data = bioann,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
 n=321 (202 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.49839171      0 1.0000000 1.0035917 0.06996966
2 0.04331689      1 0.5016083 0.8868468 0.06725061

Node number 1: 321 observations,    complexity param=0.4983917
  mean=41.79128, MSE=48.14335
  left son=2 (116 obs) right son=3 (73 obs), 132 observations remain
  Primary splits:
      TN  < 1.1555 to the right, improve=0.09447965, (132 missing)
      HQI < 18.125 to the left,  improve=0.07041631, (7 missing)
      TP  < 0.0975 to the right, improve=0.05988583, (76 missing)

Node number 2: 116 observations
  mean=38.1681, MSE=34.29717

Node number 3: 73 observations
  mean=43.87671, MSE=51.69028
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%    25%   50%    75%   95%
[1,] 18.125 0.07148321  38.64198   42.89700 0.001 17.75 18.125 18.25 20.125 23.25
```

ANALYSIS: FISH IBI VS. TP IN THE ANNUAL DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%     25%    50%    75%    95%
[1,] 0.0975 0.08040092  43.37895      39.39 0.001 0.0755 0.09325 0.0975 0.0975 1.7675
```

ANALYSIS: FISH IBI VS. TN IN THE ANNUAL DATASET (nCPA)

```
         cp        r2 mean left mean right pperm      5%  25%  50%    75%    95%
[1,] 1.1555 0.1584995  43.87671    38.1681 0.001 1.02725 1.04 1.11 1.1585 1.6505
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = FISHIBI ~ TP + TN + HQI, data = bioind,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=321 (42 observations deleted due to missingness)

        CP nsplit rel error    xerror       xstd
1 0.4802884      0 1.0000000 1.0058967 0.07020171
2 0.0455718      1 0.5197116 0.9427023 0.06879063

Node number 1: 321 observations,    complexity param=0.4802884
  mean=41.77882, MSE=48.17382
  left son=2 (105 obs) right son=3 (81 obs), 135 observations remain
  Primary splits:
      TN  < 1.195  to the right, improve=0.07509676, (135 missing)
      HQI < 18.125 to the left,  improve=0.07136043, (7 missing)
      TP  < 0.0955 to the right, improve=0.04921110, (79 missing)

Node number 2: 105 observations
  mean=38.13333, MSE=36.52032

Node number 3: 81 observations
  mean=43.17284, MSE=51.87753
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%    25%    50%    75%   95%
[1,] 18.125 0.07236082  38.61585   42.88362 0.001 16.875 18.125 18.375 20.125 23.25
```

ANALYSIS: FISH IBI VS. TP IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm      5%    25%     50%     75%     95%
[1,] 0.0955 0.06586289  43.31395   39.60897 0.003 0.09025 0.0955 0.10125 0.11525 0.19875
```

ANALYSIS: FISH IBI VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm      5%      25%   50%      75%   95%
[1,] 1.195 0.1262537  43.17284   38.13333 0.002 0.65975 1.039725 1.195 1.496625 4.791
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = FISHIBI ~ TP + TN + HQI, data = biocrit,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
 n=236 (57 observations deleted due to missingness)

        CP nsplit rel error    xerror       xstd
1 0.2011587      0 1.0000000 1.0075931 0.08297604
2 0.1756756      2 0.5976826 0.9581565 0.08047993

Node number 1: 236 observations,    complexity param=0.2011587
  mean=41.90678, MSE=53.59724
  left son=2 (134 obs) right son=3 (67 obs), 35 observations remain
  Primary splits:
      HQI < 21.25   to the left,  improve=0.07979917, (35 missing)
      TP  < 0.0975  to the right, improve=0.04983256, (66 missing)
      TN  < 4.8245  to the right, improve=0.03931795, (118 missing)

Node number 2: 134 observations,    complexity param=0.2011587
  mean=40.14179, MSE=51.06572
  left son=4 (67 obs) right son=5 (30 obs), 37 observations remain
  Primary splits:
      TP  < 0.0975  to the right, improve=0.08860471, (37 missing)
      HQI < 16.75   to the left,  improve=0.06185457, (0 missing)
      TN  < 6.33625 to the right, improve=0.04933744, (66 missing)

Node number 3: 67 observations
  mean=44.89552, MSE=49.13834

Node number 4: 67 observations
  mean=37.00746, MSE=43.71263

Node number 5: 30 observations
  mean=42.41667, MSE=44.6347
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%   25%   50%   75%   95%
[1,] 21.25 0.09057204  40.14179   44.89552 0.001 15.375 16.75 21.25 21.25 22.75
```

ANALYSIS: FISH IBI VS. TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%     25%    50%    75%   95%
[1,] 0.0975 0.07158481  43.27679   39.17982 0.012 0.064 0.09025 0.0975 0.1125 2.065
```

ANALYSIS: FISH IBI VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
       cp       r2 mean left mean right pperm     5%  25%     50%   75%     95%
[1,] 4.8245 0.0774558  40.71717   35.13158 0.026 0.6407 1.18 2.53635 5.125 6.44225
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN PAIRED OBSERVATIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = FISHIBI ~ TP + TN + HQI, data = biomatch,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=438 (87 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.16425245      0 1.0000000 1.0051850 0.05934672
2 0.07345094      4 0.3429902 0.9759769 0.05879448

Node number 1: 438 observations,    complexity param=0.1642525
  mean=42.21233, MSE=51.53254
  left son=2 (118 obs) right son=3 (295 obs), 25 observations remain
  Primary splits:
     HQI < 18.25  to the left,  improve=0.06137855, (25 missing)
     TP  < 2.03   to the left,  improve=0.01391925, (253 missing)

Node number 2: 118 observations,    complexity param=0.1642525
  mean=39.25424, MSE=51.61333
  left son=4 (10 obs) right son=5 (32 obs), 76 observations remain
  Primary splits:
     TP  < 0.381  to the left,  improve=0.03053920, (73 missing)
     HQI < 16.75  to the left,  improve=0.02550893, (0 missing)
     TN  < 0.965  to the left,  improve=0.01911753, (85 missing)

Node number 3: 295 observations,    complexity param=0.1642525
  mean=43.30847, MSE=47.39976
  left son=6 (224 obs) right son=7 (71 obs)
  Primary splits:
     HQI < 23.25  to the left,  improve=0.02457101, (0 missing)
     TN  < 0.9135 to the right, improve=0.01311013, (221 missing)
     TP  < 2.03   to the left,  improve=0.01033969, (176 missing)
```

```
Node number 4: 10 observations
  mean=30.6, MSE=24.64

Node number 5: 32 observations
  mean=38.65625, MSE=44.03809

Node number 6: 224 observations,    complexity param=0.1642525
  mean=42.70089, MSE=46.12036
  left son=12 (11 obs) right son=13 (46 obs), 167 observations remain
  Primary splits:
      TP  < 0.095  to the right, improve=0.02246379, (132 missing)
      TN  < 0.9135 to the right, improve=0.01065100, (167 missing)

Node number 7: 71 observations
  mean=45.22535, MSE=46.5971

Node number 12: 11 observations
  mean=36.45455, MSE=59.70248

Node number 13: 46 observations
  mean=43.02174, MSE=46.10822
```

ANALYSIS: FISH IBI VS. HABITAT (HQI) IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%   25%   50%   75%    95%
[1,] 18.25 0.06456082  39.25424   43.30847 0.001 16.75 18.25 19.25 20.25 22.275
```

ANALYSIS: FISH IBI VS. TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
      cp         r2 mean left mean right pperm    5%      25%   50%  75%   95%
[1,] 2.03 0.03657988  39.54762   44.05882 0.119 0.025 0.093625 1.735 2.03 3.295
```

ANALYSIS: FISH IBI VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%    25%    50%  75%   95%
[1,] 6.4895 0.02726080  38.76238   35.28571 0.554 0.5215 0.8885 1.1535 6.08 9.087
```

ANALYSIS: RBIBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = RBIBI ~ TP + TN + HQI, data = bioall,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=145 (377 observations deleted due to missingness)
```

```
        CP nsplit rel error   xerror      xstd
1 0.1774763      0 1.0000000 1.002542 0.1360853
2 0.1720651      2 0.6450475 1.018992 0.1520750

Node number 1: 145 observations,    complexity param=0.1774763
  mean=27.89621, MSE=35.16752
  left son=2 (95 obs) right son=3 (49 obs), 1 observation remains
  Primary splits:
      HQI < 20.75    to the left,  improve=0.14822400, (1 missing)
      TP  < 0.105    to the right, improve=0.05430776, (40 missing)
      TN  < 2.1275   to the right, improve=0.03202169, (64 missing)

Node number 2: 95 observations
  mean=26.22579, MSE=26.81641

Node number 3: 49 observations,    complexity param=0.1774763
  mean=31.06122, MSE=36.38401
  left son=6 (10 obs) right son=7 (20 obs), 19 observations remain
  Primary splits:
      TP  < 0.0725   to the right, improve=0.10897690, (15 missing)
      TN  < 0.951525 to the right, improve=0.04232621, (26 missing)

Node number 6: 10 observations
  mean=28.05, MSE=58.8225

Node number 7: 20 observations
  mean=33.5, MSE=7.675
```

ANALYSIS: RBIBI VS. HABITAT (HQI) IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%   25%    50%    75%    95%
[1,] 20.75 0.2101721  26.22579   31.80851 0.001 19.875 20.75 21.125 21.375 22.375
```

ANALYSIS: RBIBI VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm      5%     25%   50%    75%   95%
[1,] 0.105 0.07495033    29.879   26.62727 0.051 0.05125 0.07125 0.105 0.1125 0.167
```

ANALYSIS: RBIBI VS. TN IN THE PERIOD OF RECORD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%      25%  50%    75%   95%
[1,] 2.1275 0.06516088  29.53305   26.34091 0.193 0.585 0.843875 1.03 2.1275 2.345
```

ANALYSIS: RBIBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = RBIBI ~ TP + TN + HQI, data = bioann,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
 n=237 (286 observations deleted due to missingness)

        CP nsplit rel error    xerror       xstd
1 0.1848283      0 1.0000000 1.0147215 0.09773247
2 0.1280303      2 0.6303434 0.9319892 0.10775869

Node number 1: 237 observations,    complexity param=0.1848283
  mean=28.40063, MSE=36.40849
  left son=2 (158 obs) right son=3 (75 obs), 4 observations remain
  Primary splits:
      HQI < 21.125  to the left,  improve=0.154305700, (4 missing)
      TP  < 0.08075 to the right, improve=0.043493310, (73 missing)
      TN  < 0.87475 to the right, improve=0.009426222, (128 missing)

Node number 2: 158 observations,    complexity param=0.1848283
  mean=26.74335, MSE=29.42782
  left son=4 (10 obs) right son=5 (89 obs), 59 observations remain
  Primary splits:
      TP  < 0.0775  to the right, improve=0.03437085, (44 missing)
      HQI < 18.875  to the left,  improve=0.02747622, (0 missing)
      TN  < 0.7315  to the right, improve=0.01528405, (77 missing)

Node number 3: 75 observations
  mean=31.86, MSE=34.7904

Node number 4: 10 observations
  mean=21.85, MSE=22.8025

Node number 5: 89 observations
  mean=28.02191, MSE=29.23382
```

ANALYSIS: RBIBI VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
         cp        r2 mean left mean right pperm     5%    25%    50%    75%    95%
[1,] 21.125 0.1549966  26.74335      31.86 0.001 19.625 20.625 21.125 21.125 22.125
```

ANALYSIS: RBIBI VS. TP IN THE ANNUAL DATASET (nCPA)

```
          cp         r2 mean left mean right pperm    5%    25%     50%    75%    95%
[1,] 0.08075 0.06730738  30.39255   27.04701 0.015 0.065 0.0775 0.08075 0.0895 1.7675
```

ANALYSIS: RBIBI VS. TN IN THE ANNUAL DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%   25%    50%    75%    95%
[1,] 0.87475 0.02499873  29.77794   27.91333 0.572 0.5533 0.655 0.8625 1.0625 1.7308
```

ANALYSIS: RBIBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = RBIBI ~ TP + TN + HQI, data = bioind,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=235 (128 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.2050663      0 1.0000000 1.011515 0.09812207
2 0.1561291      2 0.5898674 1.016324 0.12655537

Node number 1: 235 observations,    complexity param=0.2050663
  mean=28.42319, MSE=36.15539
  left son=2 (157 obs) right son=3 (74 obs), 4 observations remain
  Primary splits:
     HQI < 21.125  to the left,  improve=0.14383780, (4 missing)
     TP  < 0.0805  to the right, improve=0.03276337, (76 missing)
     TN  < 0.7575  to the right, improve=0.01416640, (129 missing)

Node number 2: 157 observations,    complexity param=0.2050663
  mean=26.83408, MSE=29.79892
  left son=4 (21 obs) right son=5 (70 obs), 66 observations remain
  Primary splits:
     TP  < 0.0805  to the right, improve=0.02927310, (46 missing)
     HQI < 19.625  to the left,  improve=0.02727256, (0 missing)
     TN  < 0.72925 to the right, improve=0.01542270, (79 missing)

Node number 3: 74 observations
  mean=31.76351, MSE=34.56232

Node number 4: 21 observations
  mean=25.09524, MSE=20.53855

Node number 5: 70 observations
  mean=28.585, MSE=28.89852
```

ANALYSIS: RBIBI VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
        cp       r2 mean left mean right pperm     5%    25%    50%    75%   95%
[1,] 21.125 0.1444901  26.83408   31.76351 0.001 19.625 20.375 21.125 21.125 21.75
```

ANALYSIS: RBIBI VS. TP IN THE INDEX PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm      5%    25%    50%    75%  95%
[1,] 0.0805 0.05101748  30.39615   27.32083 0.051 0.05785 0.0805 0.0975 0.1175 1.73
```

ANALYSIS: RBIBI VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%    25%    50%    75%   95%
[1,] 0.7575 0.03772087  30.64375   28.09756 0.381 0.5335 0.6625 0.7575 0.8725 1.645
```

ANALYSIS: RBIBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = RBIBI ~ TP + TN + HQI, data = biocrit,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=195 (98 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.2907306      0 1.0000000 1.008655 0.1105318
2 0.1148645      2 0.4185388 1.008284 0.1305455

Node number 1: 195 observations,    complexity param=0.2907306
  mean=28.72026, MSE=40.56381
  left son=2 (117 obs) right son=3 (44 obs), 34 observations remain
  Primary splits:
      HQI < 21.25   to the left,  improve=0.15058930, (34 missing)
      TP  < 0.065   to the right, improve=0.03593417, (64 missing)
      TN  < 0.605   to the right, improve=0.01585912, (111 missing)

Node number 2: 117 observations,    complexity param=0.2907306
  mean=26.72607, MSE=35.58097
  left son=4 (32 obs) right son=5 (20 obs), 65 observations remain
  Primary splits:
      TN  < 1.355   to the left,  improve=0.05687192, (65 missing)
      HQI < 18.25   to the left,  improve=0.05400243, (0 missing)
      TP  < 0.98    to the left,  improve=0.03445935, (37 missing)

Node number 3: 44 observations
  mean=32.82955, MSE=39.79481
```

```
Node number 4: 32 observations
  mean=24.88906, MSE=35.2159

Node number 5: 20 observations
  mean=29.275, MSE=21.63688
```

ANALYSIS: RBIBI VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 21.25 0.1676476  26.72607   32.82955 0.001 18.25 19.75 21.25 21.25 22.75
```

ANALYSIS: RBIBI VS. TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 0.065 0.05616164  30.97885   27.28571 0.099 0.055 0.065 0.065 0.085 0.966
```

ANALYSIS: RBIBI VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%   25%     50%    75%   95%
[1,] 0.605 0.04391234  31.30769   27.92887 0.406 0.5545 0.605 0.91775 1.2435 1.575
```

ANALYSIS: RBIBI VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PAIRED OBSERVATIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = RBIBI ~ TP + TN + HQI, data = biomatch,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=357 (168 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.31204210      0 1.0000000 1.0071193 0.07391566
2 0.02376876      2 0.3759158 0.9179964 0.07922679

Node number 1: 357 observations,    complexity param=0.3120421
  mean=28.56989, MSE=40.6321
  left son=2 (231 obs) right son=3 (104 obs), 22 observations remain
  Primary splits:
      HQI < 21.25 to the left,  improve=0.13284210, (22 missing)
      TP  < 0.055 to the right, improve=0.01831483, (242 missing)

Node number 2: 231 observations,    complexity param=0.3120421
  mean=26.94134, MSE=33.18921
  left son=4 (32 obs) right son=5 (14 obs), 185 observations remain
```

```
  Primary splits:
    TN  < 1.713 to the left,  improve=0.04005650, (185 missing)
    HQI < 18.25 to the left,  improve=0.03723753, (0 missing)

Node number 3: 104 observations
  mean=32.125, MSE=41.95553

Node number 4: 32 observations
  mean=23.17031, MSE=27.0992

Node number 5: 14 observations
  mean=28.78571, MSE=15.88265
```

ANALYSIS: RBIBI VS. HABITAT (HQI) IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 21.25 0.1380638  26.94134     32.125 0.001 19.75 20.25 21.25 21.25 21.75
```

ANALYSIS: RBIBI VS.TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%   25%   50%   75% 95%
[1,] 0.055 0.06405695  31.64286   26.99455 0.083 0.055 0.055 0.065 0.065 1.1
```

ANALYSIS: RBIBI VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%   25%   50%   75%  95%
[1,] 1.713 0.09626322  24.62073   28.52941 0.123 0.725 0.935 1.628 1.713 2.24
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + HQI, data = bioall, 
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=135 (387 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.1972478      0 1.0000000 1.017211 0.1833763
2 0.1140018      1 0.8027522 1.033759 0.1648756

Node number 1: 135 observations,    complexity param=0.1972478
  mean=0.5256, MSE=0.110368 
  left son=2 (75 obs) right son=3 (51 obs), 9 observations remain
  Primary splits:
    TP  < 0.08925 to the right, improve=0.15360240, (9 missing)
```

```
    HQI < 19.625  to the left,  improve=0.10859890, (10 missing)

Node number 2: 75 observations
  mean=0.4134667, MSE=0.05455532

Node number 3: 51 observations
  mean=0.6880392, MSE=0.1542962
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%    25%    50%   75%   95%
[1,] 19.625 0.1178750 0.4110758   0.638983 0.004 17.75 18.375 19.625 20.75 22.75
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
         cp        r2 mean left mean right pperm     5%   25%     50%   75%    95%
[1,] 0.08925 0.1606124 0.6880392  0.4134667 0.002 0.05325 0.068 0.08925 0.124 0.2455
```

ANALYSIS: SECCHI TRANSPARENCY VS.TN IN THE PERIOD OF RECORD DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%       25%   50%   75%     95%
[1,] 2.559 0.02575384 0.5630137  0.4523437 0.609 0.615 0.8149688 1.582 2.575 6.70225
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + HQI, data = bioann,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=248 (275 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.20204119      0 1.0000000 1.0091001 0.07908005
2 0.09245854      1 0.7979588 0.8799153 0.07774324

Node number 1: 248 observations,    complexity param=0.2020412
  mean=0.5288206, MSE=0.1028403
  left son=2 (154 obs) right son=3 (82 obs), 12 observations remain
  Primary splits:
    TP  < 0.077   to the right, improve=0.11796200, (12 missing)
    HQI < 18.375  to the left,  improve=0.06605772, (19 missing)
    TN  < 0.71225 to the right, improve=0.03693996, (45 missing)

Node number 2: 154 observations
  mean=0.4322403, MSE=0.07495456
```

```
Node number 3: 82 observations
  mean=0.6693598, MSE=0.1074201
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%   25%    50%    75%      95%
[1,] 18.375 0.07262926 0.3798413  0.5719127 0.001 17.75 18.25 18.375 19.625 20.96875
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE ANNUAL DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%    25%   50%   75%     95%
[1,] 0.077 0.1287906 0.6693598  0.4322403 0.001 0.048 0.0565 0.077 0.078 0.26475
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE ANNUAL DATASET (nCPA)

```
          cp        r2 mean left mean right pperm   5%     25%   50%      75%    95%
[1,] 0.71225 0.0501683 0.6847414  0.4900575  0.03 0.58 0.70225 0.715 1.584938 4.9909
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + HQI, data = bioind,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=247 (116 observations deleted due to missingness)

         CP nsplit rel error   xerror       xstd
1 0.1894054      0 1.0000000 1.006090 0.08937981
2 0.1064514      1 0.8105946 0.925228 0.09216917

Node number 1: 247 observations,    complexity param=0.1894054
  mean=0.4966194, MSE=0.1011741
  left son=2 (164 obs) right son=3 (69 obs), 14 observations remain
  Primary splits:
     TP  < 0.0695  to the right, improve=0.12128090, (14 missing)
     HQI < 18.375  to the left,  improve=0.06029548, (21 missing)
     TN  < 0.6335  to the right, improve=0.03452808, (49 missing)

Node number 2: 164 observations
  mean=0.4092835, MSE=0.0714134
```

```
Node number 3: 69 observations
  mean=0.6590942, MSE=0.1238402
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%    25%    50%    75%    95%
[1,] 18.375 0.06679109  0.358373  0.5404755 0.001 17.75 18.125 18.375 20.375 23.125
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE INDEX PERIOD DATASET (nCPA)

```
         cp        r2 mean left mean right pperm      5%    25%    50%      75%    95%
[1,] 0.0695 0.1301471 0.6590942  0.4092835 0.001 0.03675 0.0605 0.0695 0.0695625 0.1485
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%     25%   50%      75%    95%
[1,] 0.6335 0.04551924 0.6785714  0.4641808 0.045 0.502 0.60625 0.636 0.941875 5.0685
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + HQI, data = biocrit,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=182 (111 observations deleted due to missingness)

         CP nsplit rel error    xerror       xstd
1 0.2990103      0 1.0000000 1.0071325 0.11022095
2 0.1591292      1 0.7009897 0.8492445 0.08941356

Node number 1: 182 observations,    complexity param=0.2990103
  mean=0.513489, MSE=0.1096731
  left son=2 (118 obs) right son=3 (43 obs), 21 observations remain
  Primary splits:
      TP  < 0.069     to the right, improve=0.13225940, (21 missing)
      TN  < 0.5575    to the right, improve=0.04907874, (51 missing)
      HQI < 18.25     to the left,  improve=0.03536519, (49 missing)

Node number 2: 118 observations
  mean=0.4067373, MSE=0.06397431

Node number 3: 43 observations
  mean=0.6961628, MSE=0.1498405
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
      cp         r2 mean left mean right pperm   5%   25%   50%   75%   95%
[1,] 18.25 0.04883256    0.4275  0.5842308 0.081 16.5 17.75 18.25 20.25 21.25
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
      cp        r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 0.069 0.1587273 0.6961628  0.4067373 0.001 0.055 0.065 0.069 0.089 0.225
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
       cp         r2 mean left mean right pperm       5%    25%   50%   75%     95%
[1,] 0.5575 0.07761649 0.7367857  0.4568803 0.027 0.532025 0.5575 0.655 0.763 2.76025
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PAIRED OBSERVATIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + HQI, data = biomatch,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=228 (297 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.3248921      0 1.0000000 1.004559 0.1332763
2 0.0953041      1 0.6751079 1.050879 0.1136351

Node number 1: 228 observations,    complexity param=0.3248921
  mean=0.5571491, MSE=0.1530292
  left son=2 (113 obs) right son=3 (54 obs), 61 observations remain
  Primary splits:
      TP  < 0.062  to the right, improve=0.10229650, (61 missing)
      HQI < 18.25  to the left,  improve=0.02729419, (30 missing)

Node number 2: 113 observations
  mean=0.475885, MSE=0.1127687

Node number 3: 54 observations
  mean=0.7884259, MSE=0.2002239
```

ANALYSIS: SECCHI TRANSPARENCY VS. HABITAT (HQI) IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
      cp         r2 mean left mean right pperm   5%   25%   50%   75%   95%
[1,] 18.25 0.03143942 0.4585484  0.6080882 0.116 17.5 17.75 18.25 21.75 23.25
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
      cp       r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 0.062 0.1315872 0.7884259   0.475885 0.001 0.062 0.062 0.062 0.065 0.304
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%     25%   50%   75%    95%
[1,] 0.8605 0.04272416 0.6866667  0.5273418 0.287 0.573 0.80025 0.883 3.254 9.2965
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + HQI, data = bioall,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=93 (429 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.17475205      0  1.000000 1.0115470 0.5130478
2 0.04420316      1  0.825248 0.9479896 0.4564116

Node number 1: 93 observations,    complexity param=0.1747521
  mean=6.672204, MSE=47.4522
  left son=2 (45 obs) right son=3 (39 obs), 9 observations remain
  Primary splits:
      HQI < 19.625 to the right, improve=0.14409480, (9 missing)
      TP  < 0.974  to the left,  improve=0.06115592, (0 missing)
      TN  < 3.915  to the left,  improve=0.03554306, (17 missing)

Node number 2: 45 observations
  mean=4.317333, MSE=10.526

Node number 3: 39 observations
  mean=9.834231, MSE=81.23575
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HQI IN THE PERIOD OF RECORD DATASET (nCPA)

```
       cp       r2 mean left mean right pperm   5%   25%    50%    75%   95%
[1,] 19.625 0.1486521   9.83423   4.317333 0.003 17.5 19.25 19.625 19.625 19.75
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
      cp        r2 mean left mean right pperm      5%   25%     50%   75%  95%
[1,] 0.065 0.05766784  4.151786   7.757923 0.166 0.05575 0.065 0.08925 0.808 1.06
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
      cp       r2 mean left mean right pperm     5%   25%   50%   75%  95%
[1,] 3.915 0.0372849  6.227018   9.544737 0.538 0.6856 1.322 3.879 4.473 7.47
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + HQI, data = bioann,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=135 (388 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.08319319      0 1.0000000 1.012767 0.8896081
2 0.01153332      1 0.9168068 1.299459 0.8880091

Node number 1: 135 observations,    complexity param=0.08319319
  mean=6.342635, MSE=265.3255
  left son=2 (82 obs) right son=3 (19 obs), 34 observations remain
  Primary splits:
      HQI < 18.25   to the right, improve=0.05939647, (12 missing)
      TP  < 0.631   to the right, improve=0.01671782, (0 missing)

Node number 2: 82 observations
  mean=5.06196, MSE=27.55846

Node number 3: 19 observations
  mean=17.49211, MSE=1609.435
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
cp        r2 mean left mean right pperm 5%  25%   50%  75%  95%
[1,] 18.25 0.05971362  13.49833   4.111674 0.133 15 17.5 18.25 18.5 18.5
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE ANNUAL DATASET (nCPA)

```
      cp        r2 mean left mean right pperm       5%   25%     50%   75%  95%
[1,] 0.631 0.01671782  7.733571   3.153659 0.593 0.0479625 0.146 0.43325 0.631 1.38
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE ANNUAL DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%   25%  50%  75%   95%
[1,] 1.2825 0.01251725  5.022288   9.235965 0.918 1.2825 1.498 1.59 2.24 6.986
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + HQI, data = bioind,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=125 (238 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.20192172      0 1.0000000 1.010021 0.4108313
2 0.05804111      1 0.7980783 1.598334 0.4182301

Node number 1: 125 observations,    complexity param=0.2019217
  mean=5.600538, MSE=55.26345
  left son=2 (79 obs) right son=3 (12 obs), 34 observations remain
  Primary splits:
      TP  < 1.5075 to the right, improve=0.05474723, (0 missing)
      TN  < 6.5955 to the right, improve=0.02493820, (34 missing)

Node number 2: 79 observations
  mean=5.31832, MSE=39.66063

Node number 3: 12 observations
  mean=14.75375, MSE=198.3233
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%      25%   50%   75%   95%
[1,] 16.375 0.04155535  8.437143   4.743473 0.182 16.375 16.46875 17.75 19.75 23.25
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%   25%     50%     75%    95%
[1,] 1.5075 0.05474723  6.336956   1.492105 0.104 0.0465 0.277 0.47925 1.25525 1.5075
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm      5%      25%     50%    75%    95%
[1,] 6.5955 0.02674805   7.12426   3.192308 0.593 0.66725 1.460313 1.90725 5.8665 7.1185
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + + HQI, data = biocrit,
```

```
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=90 (203 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.3053869      0  1.000000 1.022469 0.3699820
2 0.0333487      1  0.694613 1.476002 0.3947486

Node number 1: 90 observations,    complexity param=0.3053869
  mean=7.316333, MSE=92.80755
  left son=2 (60 obs) right son=3 (11 obs), 19 observations remain
  Primary splits:
      HQI < 16.5    to the right, improve=0.14158690, (19 missing)
      TP  < 2.3375  to the right, improve=0.07165494, (0 missing)
      TN  < 7.92975 to the right, improve=0.06593289, (31 missing)

Node number 2: 60 observations
  mean=5.066167, MSE=30.4653

Node number 3: 11 observations
  mean=16.34545, MSE=361.2693
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
       cp       r2 mean left mean right pperm   5%  25%  50%   75%   95%
[1,] 16.5 0.1693218  16.34545   5.066167 0.008 15.5 16.5 16.5 18.25 20.75
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%  25%    50%    75%    95%
[1,] 2.3375 0.07165494    8.4696       1.55 0.136 0.313 1.15 1.5275 2.3375 2.3375
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
       cp        r2 mean left mean right pperm     5%    25%    50%   75%   95%
[1,] 4.14 0.05720832  9.234634      8.675  0.33 0.5695 1.0875 1.3585 7.371 8.181
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PAIRED OBSERVATIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + HQI, data = biomatch,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=63 (462 observations deleted due to missingness)
```

```
        CP nsplit rel error   xerror      xstd
1 0.23642567      0 1.0000000 1.025973 0.5103646
2 0.04165734      1 0.7635743 1.022708 0.3834683

Node number 1: 63 observations,    complexity param=0.2364257
  mean=7.778571, MSE=137.8207
  left son=2 (47 obs) right son=3 (11 obs), 5 observations remain
  Primary splits:
      HQI < 17.5   to the right, improve=0.22569680, (5 missing)
      TP  < 0.0855 to the left,  improve=0.03877165, (7 missing)
      TN  < 2.316  to the right, improve=0.02058742, (27 missing)

Node number 2: 47 observations
  mean=5.145532, MSE=30.35383

Node number 3: 11 observations
  mean=19.97273, MSE=473.0238
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. HQI IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
     cp        r2 mean left mean right pperm    5% 25% 50% 75%   95%
[1,] 17 0.3529157  19.97273   5.145532 0.001 16.25  17  17  18 21.25
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%    25%    50%   75%  95%
[1,] 0.0855 0.07087514  2.439231   8.246512 0.262 0.0605 0.0855 0.1555 1.182 2.49
```

ANALYSIS: SPECTROPHOTOMETRIC CHL-A VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%   25%   50%   75%    95%
[1,] 0.7155 0.04393775     3.851        8.7 0.661 0.623 0.779 1.301 1.666 2.316
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TP + TN + HQI, data = bioall, 
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=90 (432 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.12868778      0 1.0000000 1.018915 0.6226745
2 0.01232473      2 0.7426244 1.159953 0.5405223
```

```
Node number 1: 90 observations,    complexity param=0.1286878
  mean=6.906611, MSE=77.73436
  left son=2 (38 obs) right son=3 (34 obs), 18 observations remain
  Primary splits:
      TN  < 1.335  to the left,  improve=0.04668549, (17 missing)
      TP  < 0.38   to the right, improve=0.03059765, (3 missing)
      HQI < 18.375 to the left,  improve=0.02341658, (4 missing)

Node number 2: 38 observations
  mean=4.308553, MSE=20.88802

Node number 3: 34 observations,    complexity param=0.1286878
  mean=9.936324, MSE=156.7199
  left son=6 (13 obs) right son=7 (17 obs), 4 observations remain
  Primary splits:
      TN  < 1.1325 to the left,  improve=0.13259910, (4 missing)
      TP  < 0.1225 to the left,  improve=0.12899160, (0 missing)
      HQI < 19.25  to the left,  improve=0.08724583, (2 missing)

Node number 6: 13 observations
  mean=4.831538, MSE=8.384705

Node number 7: 17 observations
  mean=14.625, MSE=252.5131
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm    5%    25%   50%   75%   95%
[1,] 19.25 0.02790208  5.520714    8.53869 0.565 17.625 18.375 19.25 19.25 20.25
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
       cp         r2 mean left mean right pperm    5%    25%    50%  75%     95%
[1,] 0.38 0.03080267  8.008333    4.49875 0.457 0.065 0.1225 0.2095 0.38 0.48875
```

ANALYSIS: FLUOROMETRIC CHL-A VS.TN IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp         r2 mean left mean right pperm    5%    25%   50%    75%   95%
[1,] 1.335 0.05351579  5.176625   9.426667 0.312 0.965 1.1325 1.335 2.2475 3.915
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TP + TN + HQI, data = bioann,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
```

```
 n=110 (413 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.5248133      0 1.0000000 1.032336 0.3426281
2 0.0765306      1 0.4751867 1.044455 0.3353817

Node number 1: 110 observations,    complexity param=0.5248133
  mean=10.44005, MSE=162.8896
  left son=2 (33 obs) right son=3 (44 obs), 33 observations remain
  Primary splits:
      TN  < 2.5225   to the right, improve=0.05176019, (29 missing)
      TP  < 0.0615   to the left,  improve=0.03789520, (8 missing)
      HQI < 20.25    to the right, improve=0.01268401, (8 missing)

Node number 2: 33 observations
  mean=4.871364, MSE=9.804666

Node number 3: 44 observations
  mean=15.17557, MSE=186.154
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5% 25%   50%   75%    95%
[1,] 20.25 0.01308977  11.62381    8.55218 0.863 15.125  18 19.25 20.25 22.375
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TP IN THE ANNUAL DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5%    25%   50%       75%   95%
[1,] 0.0615 0.03868564  6.614259   12.46247 0.385 0.0615 0.0615 0.185 0.3229375 1.175
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE ANNUAL DATASET (nCPA)

```
         cp         r2 mean left mean right pperm     5% 25%    50%    75%    95%
[1,] 2.5225 0.08687327  12.04523    3.54625 0.073 0.9939 1.8 2.4525 2.5225 2.7775
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TP + TN + NH4 + NOX + SRP + HQI, data = bioind,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=110 (253 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.3765427      0 1.0000000 1.018330 0.3824738
2 0.1742876      1 0.6234573 1.091528 0.3422390
```

```
Node number 1: 110 observations,    complexity param=0.3765427
  mean=11.04345, MSE=240.3756
  left son=2 (35 obs) right son=3 (41 obs), 34 observations remain
  Primary splits:
      TN  < 2.1825  to the right, improve=0.05714005, (31 missing)
      TP  < 0.19075 to the right, improve=0.04120479, (10 missing)

Node number 2: 35 observations
  mean=5.663, MSE=20.89318

Node number 3: 41 observations
  mean=18.13415, MSE=384.2383
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm    5%    25%   50%   75%   95%
[1,] 18.875 0.01450581  9.208333   13.09920  0.79 17.25 18.875 20.25 22.25 22.75
```

ANALYSIS: FLUOROMETRIC CHL-A VS.TP IN THE INDEX PERIOD DATASET (nCPA)

```
          cp         r2 mean left mean right pperm     5%    25%     50%     75%   95%
[1,] 0.19075 0.04223754  14.33802   7.501351 0.243 0.0625 0.1225 0.19075 0.19375 1.195
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
         cp         r2 mean left mean right pperm       5%  25%   50%    75%    95%
[1,] 2.1825 0.07814729  14.47712    4.41975 0.134 1.027225 1.11 1.305 2.1825 2.2425
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
mvpart(form = CHLAFLUORO ~ TP + TN + HQI, data = biocrit,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=87 (206 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.75098050      0 1.0000000 1.0127449 0.6502199
2 0.07296209      1 0.2490195 0.9742743 0.6474421

Node number 1: 87 observations,    complexity param=0.7509805
  mean=11.13069, MSE=383.3476
  left son=2 (47 obs) right son=3 (25 obs), 15 observations remain
  Primary splits:
      TN  < 2.18135 to the right, improve=0.01939596, (31 missing)
```

```
    TP  < 0.075   to the left,  improve=0.01046382, (14 missing)

Node number 2: 47 observations
  mean=4.837872, MSE=13.37752

Node number 3: 25 observations
  mean=18.2276, MSE=307.0547
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm   5%  25%   50%  75%   95%
[1,] 21.25 0.02666039  10.83723       5.6 0.677 15.5 18.5 18.75 20.5 21.25
```

ANALYSIS: FLUOROMETRIC CHL-A VS.TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm    5%   25%   50%   75%  95%
[1,] 0.065 0.01597189  8.938889   13.75218 0.951 0.073 0.165 0.275 0.598 1.23
```

ALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
          cp        r2 mean left mean right pperm    5%    25%   50%     75%  95%
[1,] 2.18135 0.06810871  12.21762   4.368571 0.306 0.615 1.2435 1.555 2.18135 2.35
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PAIRED OBSERVATTIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TP + TN + HQI, data = biomatch,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=101 (424 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.69753348      0 1.0000000 1.023982 0.5540141
2 0.06438131      1 0.3024665 1.211992 0.5509049
3 0.01000000      2 0.2380852 1.348686 0.4071719

Node number 1: 101 observations,    complexity param=0.6975335
  mean=11.92446, MSE=388.0827
  left son=2 (23 obs) right son=3 (29 obs), 49 observations remain
  Primary splits:
    TP  < 0.095  to the left,  improve=0.03049927, (14 missing)
    TN  < 2.095  to the right, improve=0.02802303, (49 missing)

Node number 2: 23 observations
  mean=5.534783, MSE=10.87765
```

```
Node number 3: 29 observations,    complexity param=0.06438131
  mean=21.09, MSE=400.1861
  left son=6 (10 obs) right son=7 (15 obs), 4 observations remain
  Primary splits:
      HQI < 17     to the left,  improve=0.08778570, (7 missing)
      TP  < 0.16   to the right, improve=0.08587671, (0 missing)

Node number 6: 10 observations
  mean=11.75, MSE=390.9105

Node number 7: 15 observations
  mean=25.6, MSE=344.852
```

ANALYSIS: FLUOROMETRIC CHL-A VS. HABITAT (HQI) IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm   5%  25%   50%   75%  95%
[1,] 20.25 0.02048334  12.32404     7.5295 0.681 15.5 18.5 19.25 20.25 20.5
```

ANALYSIS: FLUOROMETRIC CHL-A VS.TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
       cp        r2 mean left mean right pperm    5%   25%    50%    75%  95%
[1,] 0.095 0.03137172  8.063548   15.80375 0.444 0.065 0.095 0.2205 0.3295 1.46
```

ANALYSIS: FLUOROMETRIC CHL-A VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
     cp       r2 mean left mean right pperm  5%   25%   50%   75%   95%
[1,] 1.4 0.0755085  12.81861   8.445625 0.191 0.9 1.088 1.563 2.095 2.326
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PERIOD OF RECORD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TP + TN + HQI, data = bioall,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=118 (404 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.34504289      0 1.0000000 1.0207506 0.2716110
2 0.02859869      2 0.3099142 0.9759939 0.2251499

Node number 1: 118 observations,    complexity param=0.3450429
  mean=2.335508, MSE=3.490976
  left son=2 (43 obs) right son=3 (68 obs), 7 observations remain
  Primary splits:
```

```
    HQI < 19.25   to the left,  improve=0.05999548, (7 missing)
    TN  < 1.22    to the left,  improve=0.02259889, (42 missing)
    TP  < 0.07125 to the right, improve=0.01726789, (30 missing)

Node number 2: 43 observations
  mean=1.691395, MSE=0.9332899

Node number 3: 68 observations,    complexity param=0.3450429
  mean=2.66, MSE=4.045752
  left son=6 (23 obs) right son=7 (19 obs), 26 observations remain
  Primary splits:
    TN  < 1.22    to the left,  improve=0.032732440, (26 missing)
    HQI < 23.625  to the right, improve=0.017016250, (0 missing)
    TP  < 0.27175 to the left,  improve=0.009934154, (21 missing)

Node number 6: 23 observations
  mean=2.045217, MSE=1.533927

Node number 7: 19 observations
  mean=2.975526, MSE=2.750147
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) IN THE PERIOD OF RECORD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%   25%   50%    75%    95%
[1,] 19.25 0.0726982  1.691395       2.66 0.054 18.625 19.25 19.25 19.625 21.125
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE PERIOD OF RECORD DATASET (nCPA)

```
          cp         r2 mean left mean right pperm      5%     25%     50%  75%  95%
[1,] 0.07125 0.02644797  2.671379   2.066525 0.647 0.05575 0.07125 0.15175 0.38 0.82
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE PERIOD OF RECORD DATASET (nCPA)

```
       cp         r2 mean left mean right pperm      5%  25%    50%   75%    95%
[1,] 1.22 0.04534564  1.908056      2.609 0.416 0.68428 1.22 1.2925 2.281 3.6774
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE ANNUAL BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TP + TN + HQI, data = bioann,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=236 (287 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.2014766      0 1.0000000 1.008626 0.1773251
```

```
2 0.1103712      1 0.7985234 1.203375 0.1785669

Node number 1: 236 observations,    complexity param=0.2014766
  mean=2.71852, MSE=4.950706
  left son=2 (163 obs) right son=3 (19 obs), 54 observations remain
  Primary splits:
      TP  < 0.048  to the right, improve=0.04826815, (54 missing)
      TN  < 4.921  to the right, improve=0.02562002, (76 missing)
      HQI < 18.75  to the left,  improve=0.01915339, (18 missing)

Node number 2: 163 observations,    complexity param=0.1103712
  mean=2.608471, MSE=4.262156
  left son=4 (123 obs) right son=5 (23 obs), 17 observations remain
  Primary splits:
      HQI < 21.375 to the left,  improve=0.03615339, (14 missing)
      TN  < 8.083  to the right, improve=0.02588043, (21 missing)
      TP  < 0.597  to the right, improve=0.01334235, (0 missing)

Node number 3: 19 observations
  mean=4.428947, MSE=12.53877

Node number 4: 123 observations
  mean=2.430766, MSE=3.633308

Node number 5: 23 observations
  mean=3.997676, MSE=5.168725
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) IN THE ANNUAL DATASET (nCPA)

```
        cp         r2 mean left mean right pperm     5%     25%   50%    75%   95%
[1,] 18.75 0.02262645  2.084658    2.82686 0.213 15.375 18.4375 18.75 20.875 23.25
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE ANNUAL DATASET (nCPA)

```
        cp         r2 mean left mean right pperm        5%    25%   50%     75%   95%
[1,] 0.048 0.05700122  4.428947   2.608471 0.044 0.0327125 0.0465 0.048 0.06725 0.597
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE ANNUAL DATASET (nCPA)

```
        cp         r2 mean left mean right pperm      5%     25%    50%   75%    95%
[1,] 4.921 0.03353883  3.072357     1.7645 0.218 0.71125 1.46125 3.0975 4.921 6.519
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE INDEX PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TP + TN + HQI, data = bioind,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=236 (127 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.26746279      0 1.0000000 1.008864 0.1794661
2 0.03010876      2 0.4650744 1.426904 0.1628107

Node number 1: 236 observations,    complexity param=0.2674628
  mean=2.705914, MSE=4.916866
  left son=2 (152 obs) right son=3 (19 obs), 65 observations remain
  Primary splits:
      HQI < 18.75  to the left,  improve=0.02614643, (20 missing)
      TN  < 5.274  to the right, improve=0.02436293, (80 missing)
      TP  < 0.0495 to the right, improve=0.02354566, (56 missing)

Node number 2: 152 observations,    complexity param=0.2674628
  mean=2.704446, MSE=5.008579
  left son=4 (71 obs) right son=5 (21 obs), 60 observations remain
  Primary splits:
      HQI < 18.75  to the left,  improve=0.05504827, (17 missing)
      TP  < 0.0495 to the right, improve=0.05501087, (6 missing)
      TN  < 5.274  to the right, improve=0.02715089, (13 missing)


Node number 3: 19 observations
  mean=4.370789, MSE=4.685198

Node number 4: 71 observations
  mean=2.502606, MSE=2.799739

Node number 5: 21 observations
  mean=4.26, MSE=11.99348
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) IN THE INDEX PERIOD DATASET (nCPA)

```
        cp        r2 mean left mean right pperm     5%    25%   50%    75%    95%
[1,] 18.75 0.0310734   1.95753   2.834146 0.101 15.375 18.375 18.75 20.875 22.125
```

ANALYSIS: 24-HOUR DO FLUX VS.TP IN THE INDEX PERIOD DATASET (nCPA)

```
      cp       r2 mean left mean right pperm    5%   25%    50%   75%    95%
[1,] 0.0495 0.0279421  3.958333   2.659665 0.309 0.032 0.042 0.0545 0.164 0.5905
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE INDEX PERIOD DATASET (nCPA)

```
      cp       r2 mean left mean right pperm    5%   25%    50%    75%   95%
[1,] 5.274 0.03231893  3.076341   1.743889 0.278 0.622 1.261 2.0995 4.7825 7.387
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE CRITICAL PERIOD BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TP + TN + HQI, data = biocrit,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=152 (141 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.25203741      0 1.0000000 1.017174 0.2156987
2 0.04683651      1 0.7479626 1.360097 0.2021758

Node number 1: 152 observations,    complexity param=0.2520374
  mean=2.719567, MSE=4.884485
  left son=2 (95 obs) right son=3 (22 obs), 35 observations remain
  Primary splits:
     TP  < 0.349  to the right, improve=0.05877689, (34 missing)
     TN  < 1.847  to the right, improve=0.04075871, (50 missing)

Node number 2: 95 observations
  mean=2.56594, MSE=4.125429

Node number 3: 22 observations
  mean=4.319773, MSE=7.427401
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) IN THE CRITICAL PERIOD DATASET (nCPA)

```
      cp       r2 mean left mean right pperm   5%  25%  50%   75%   95%
[1,] 20.75 0.01456256  2.393333    2.89807 0.677 15.5 15.5 18.5 20.75 23.75
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE CRITICAL PERIOD DATASET (nCPA)

```
      cp      r2 mean left mean right pperm     5%    25%    50%   75%   95%
[1,] 0.349 0.0710583  3.179813   1.837059  0.05 0.05775 0.0835 0.2155 0.349 0.349
```

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE CRITICAL PERIOD DATASET (nCPA)

```
      cp         r2 mean left mean right pperm       5%    25%    50%      75%    95%
[1,] 1.847 0.05193357  3.314773      2.175 0.229 0.825275 1.428 1.847 2.54875 3.685
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) AND NUTRIENT STRESSORS (TP AND TN) IN THE PAIRED OBSERVATIONS BIOASSESSMENT DATASET (CART)

```
Call:
mvpart(form = DOFLUX ~ TP + TN + HQI, data = biomatch, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=97 (428 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.04478752      0 1.0000000 1.015134 0.2624419
2 0.03989236      1 0.9552125 1.104645 0.2795294

Node number 1: 97 observations,    complexity param=0.04478752
  mean=2.191475, MSE=3.542556
  left son=2 (38 obs) right son=3 (57 obs), 2 observations remain
  Primary splits:
      HQI < 19.75 to the left,  improve=0.03109417, (2 missing)

Node number 2: 38 observations
  mean=1.752976, MSE=2.124579

Node number 3: 57 observations
  mean=2.437544, MSE=4.342169
```

ANALYSIS: 24-HOUR DO FLUX VS. HABITAT (HQI) IN THE PAIRED OBSERVATIONS DATASET (nCPA)

```
      cp         r2 mean left mean right pperm 5%   25%   50%  75%   95%
[1,] 17.75 0.03206111  2.103889   2.177703 0.369 15 17.75 19.75 20.5 23.75
```

ANALYSIS: 24-HOUR DO FLUX VS. TP IN THE PAIRED OBSERVATIONS DATASET (nCPA)

No possible splits

ANALYSIS: 24-HOUR DO FLUX VS. TN IN THE PAIRED OBSERVATIONS DATASET (nCPA)

No possible splits

Section 2: Reservoirs

E.M. Grantz, L.B. Massey, J.T. Scott, B.E. Haggard

EXECUTIVE SUMMARY

The Clean Water Action Plan, released in 1998 by the United States Environmental Protection Agency (USEPA), established a national set of nutrient criteria for the 14 aggregate ecogregions across the United States, directing states and tribes to adopt these criteria or pursue scientifically defensible nutrient criteria at the state level. For lakes and reservoirs, the two main approaches for nutrient criteria development focus on the frequency distribution of median concentrations of a general population or select group of sites representing reference conditions and statistical analysis of stressor-response relationships between nutrients and biological response variables. Predictive approaches have focused on establishing relationships between nutrient concentrations and algae.

The objective of Section 2 was to provide statistical support to the Texas Commission on Environmental Quality (TCEQ) to aide the development of numeric nutrient criteria for Texas reservoirs by TCEQ. The first step in this process was to compile the geospatial and water quality data from 764 stations spanning 14 basins across Texas provided by TCEQ and collected under non-biased conditions. Following data reorganization and reduction, median values for each parameter were estimated at each station with 10 observations or greater and compiled into a median database. The parameters of primary concern were total phosphorus (TP), ortho-phosphate (PO_4-P; SRP), total nitrogen (TN), nitrate plus nitrite N (NO_x-N), and sestonic chlorophyll-a (chl-a). Frequency distributions including the minimum, 10th, 25th, 50th, 75th, and 90th percentiles, and maximum of these parameters were calculated for the general population at multiple spatial scales, specifically, by basin, eco-region levels III and IV, and basin by level III eco-region. Frequency distributions are presented in Section 1.1. and were intended to provide TCEQ with the percentile estimates for Texas reservoirs recommended by the USEPA for setting nutrient and chl-a criteria.

States are progressing to development of nutrient criteria, but questions remain regarding the legitimacy of promulgating one numeric criterion across areas that may contain multiple basins, various eco-regions, and a myriad of land uses. Section 2.2 provides analyses of potential geospatial variability in total nutrient (TP and TN), chl-a, and Secchi transparency for Texas reservoirs using classification and regression tree (CART) and non-parametric changepoint analysis (nCPA). Geospatial variables included land use/land cover (LULC) categories, permitted municipal wastewater treatment (WWTP) plant discharge, and regions (basin, ecoregion III, and basin by ecoregion III). Changepoints were identified in several geospatial predictor variables to describe variability in nutrient and biological response parameters. Thresholds in %Agriculture land cover were found for TN, Secchi transparency, and chl-a and ranged from 18-57% and predicted up to 20% of variability in the response variables. Thresholds were also found for TP, chl-a, and

Secchi transparency in municipal WWTP discharge both weighted and unweighted for watershed area. The area-weighted threshold for TP was in range with that found in Texas streams and rivers (0.028 vs. 0.031 mgd/km^2). Thresholds in WWTP discharge were consistently lower in magnitude for biological response variables than for TP concentrations, particularly for chl-a, suggesting that the effects of low-level WWTP discharge on Texas reservoirs may first become evident as enhanced productivity. Among the categorical region variables, grouping schemes based on basin by ecoregion III areas were most effective, particularly for TP and biological response variables. In general, CART models provided greater explanatory power for biological response than for nutrients for most geospatial categories, which likely reflects the fact that nutrient concentrations in reservoirs may have strong seasonal patterns resulting from in-lake processes rather than watershed exports.

Censored datasets present a challenge to states, tribes, and others in progressing toward statistically-based numeric criteria development. In Section 2.3, multiple approaches to estimating measures of central tendency for censored datasets were employed. These approaches included substitution of the detection limit or half the detection limit for censored values in the raw data, deletion of censored values (TP, TN, Secchi, and chl-a), or employing variations of statistical methods for calculating summary statistics in censored datasets (TP and spectrophotometric chl-a only). Using CART and nCPA, comparisons were made between metrics calculated using substitution with the detection limit and published statistical methods. Analysis of the percent difference in measures of central tendency between methods vs. the percentage of censored data indicated thresholds ranging from 16-49% censored data. Above these thresholds, the censored observations increasingly affected estimates of measures of central tendency. Because the percentage of censored values exceeded these thresholds for many stations, censoring in the Texas reservoir water quality dataset resulted in potential underestimation of the 25th percentile of station medians by 40-70% for TP and 20-30% for chl-a.

The frequency distribution approach should be used in conjunction with other statistically based methods that evaluate stressor-response relations in aquatic systems. Section 2.4 provides analyses of potential nutrient thresholds (TP and TN) for biological response (chl-a spec and Secchi transparency) for each of the 3-5 datasets with variable correction for censoring. The TP and TN thresholds identified in the analyses ranged from 0.025-0.063 mg/L TP and 0.56-0.90 mg/L TN. For TP, thresholds varied considerably between datasets with different approaches to handling censored data, especially in relationship to Secchi transparency. For TN, thresholds varied less and usually not between datasets. Results indicated that substituting half the detection limit for censored values when calculating measures of central tendency resulted in lower threshold estimates than when published statistical methods for analyzing censored datasets were used. Reasonable agreement was also found between threshold estimates when the detection limit was substituted and when published statistical methods were used. However thresholds were lower, especially for Secchi transparency, when estimates of summary statistics were made for stations with >80% censoring, indicating that censored data may be affecting changepoint analyses even when the best available methods are employed for analyzing these data.

INTRODUCTION

The Clean Water Action Plan, released in 1998, established national nutrient criteria for lakes and reservoirs in the 14 aggregate ecogregions across the United States, five of which lie partly within Texas. These numerical values were set for both causative (e.g., nutrients) and response (e.g., chlorophyll and transparency) variables which are associated with the prevention and assessment of eutrophic conditions in streams, rivers, lakes, and reservoirs. These recommendations were based on frequency distributions of found data. However, local and regional influences on water quality contribute to median concentrations and can lead to differences from US Environmental Protection Agency (USEPA) recommendations (e.g., Ice and Binkley 2003, Smith et al. 2003, Binkley et al. 2004, Evans-White et al. 2013). For example, studies have shown that nutrient levels in Pacific Northwest lakes were strongly related to typology defined by turbidity and conductivity, as well as geospatial variability (Vaga et al. 2006). Therefore, the aggregate ecoregions may be too coarse to be used for establishing nutrient criteria, and the basin or smaller ecoregion level might be more appropriate (Rohm et al. 2002). Therefore, states, tribes, and others have the option of establishing regionally-specific, scientifically defensible nutrient criteria for water bodies of various spatial scales (e.g., basins and ecoregions) specific to a regulatory jurisdiction, in lieu of adopting the criteria set by the USEPA.

Commonly applied methods to evaluating nutrient concentrations in lakes and reservoirs for the development of numeric criteria include frequency distributions and stressor-response studies. The frequency distribution method develops nutrient criteria relative to the population of water-quality data in a specific area (e.g., state, basin or ecoregion). The USEPA (2000) has suggested the 75th percentile of nutrient concentrations from reference or minimally impacted lakes and reservoir conditions as a criterion, or the 25th percentile of nutrient concentrations from a general population (i.e., all lakes and reservoirs regardless of human influence). The USEPA (2000) suggested that both approaches should result in similar criterion; however, studies have shown that a comparison between approaches can be highly variable in streams (Suplee et al. 2007, Herlihy and Sifeneos 2008) and in lakes and reservoirs (Herlihy et al. 2013). There are many additional concerns with this approach, such as limited data availability representing reference or even general populations from targeted areas. Frequency distributions from the recent National Lakes Assessment survey, which used probability-based experimental design to randomly select 1028 lakes and reservoirs across the conterminous U.S. for detailed water quality analysis, also differed from previous USEPA criteria recommendations based on found data (Herlihy et al. 2013). Furthermore, the selected percentile may not necessarily be tied to water-quality impairments. Nonetheless, the frequency distribution method is a tool that can aid states, tribes and other groups when setting nutrient criteria.

The USEPA has recommended that states and tribes use stressor-response studies to help develop nutrient criteria. In these analyses, biological conditions are evaluated over a gradient of nutrient concentrations. Classification and regression tree (CART) analysis is an empirical modeling technique that is useful for identifying ecological thresholds and hierarchical structure in predictor variables (De'ath and

Fabricius 2000). CART uses recursive partitioning to divide data into subsets that are increasingly homogeneous, invoking a tree-like classification that can explain relationships that may be difficult to reconcile with conventional linear models (Urban 2002). CART and other similar methods have been used to identify thresholds and hierarchical structure in environmental correlates of various biological processes in aquatic ecosystems (King et al. 2005, East and Sharfstein 2006). King et al. (2005) used CART to identify thresholds in nutrient concentrations which resulted in shifts in ecological structure and function. These thresholds were used to recommend specific water quality nutrient criteria for the Florida Everglades ecosystem.

States across the US are moving forward with the development of numeric nutrient criteria, although the pace varies by state and the political, legal, and environmental pressures each state is facing. Many states are concerned about the legitimacy of promulgating one numeric criterion across the whole state comprised by multiple basins, various level III and IV ecoregions, and different land uses (e.g., forest, pasture, row crop and urban). However, the development of site-specific nutrient criteria can be a costly process from the efforts needed to evaluate the physical, chemical and biological conditions of lakes and reservoirs to that required to push the numeric criteria through promulgation. For these reasons, it might not be feasible to develop numeric criteria for individual watersheds or eco-regions. However, studies have shown that almost half of the variation in nutrient concentrations can be explained by select physico-chemical properties and watershed characteristics, like runoff, elevation, land use and cover, and also eco-regions (Herlihy and Sifneos 2008). There is also evidence to suggest that undisturbed watershed conditions may not exist because of the effects of even minimal development (King and Baker 2010), atmospheric deposition (Flum and Nodvin 1995) and small catchment areas (Smith et al. 2003). Thus, states need to explore defensible approaches to aggregating reservoirs into categories to assist in the nutrient criteria development process.

Censored datasets present yet another challenge to states, tribes, and others in progressing toward statistically-based numeric nutrient criteria development because censored observations can affect analyses such as distribution fitting or threshold analysis. The true value of a censored observation is unknown, except that it falls within a range of values. Left-censored observations are bounded by zero and an analytical detection limit, and are the most common type of censored data in environmental datasets. Some environmental metrics, such as Secchi depth, can also be associated with right-censored observations. The value of right-censored observations is known only to exceed a detection limit. Common approaches for handling censored observations include deletion or substitution with either zero, the detection limit, or half the detection limit. These approaches are not statistically rigorous and can obscure existing patterns or introduce patterns to datasets that do not reflect real-world conditions.

Though less commonly employed, statistically rigorous methods for analyzing censored data do exist. Methods for calculating summary statistics, such as means, medians, standard deviations, and percentiles are well-developed (Helsel 2012). These methods extract known information, or the frequency at which censored observations occur in the dataset relative to uncensored observations. These methods can be divided into two categories: 1) parametric, or methods that require assumptions about data distribution, such as normal or lognormal, and 2) non-parametric, or methods that do not require assumptions about data distribution. Maximum likelihood estimation (MLE) and regression order statistics (ROS) analyses are both parametric methods, while the Kaplan-Meier (KM) analysis is non-parametric. As a non-parametric method, KM is generally preferred to MLE or ROS, but the utility of each method varies according to number of total observations and percentage of censored observations in a dataset.

The objectives of this chapter are:

1) to discuss the frequency distribution of median nutrient concentrations and response variables for Texas reservoirs acquired from the Texas Commission on Environmental Quality (TCEQ) at various spatial scales including individual basins, level III ecoregions, level IV ecoregions, and basin-level III ecoregion combinations;
2) to explore the relationship between median nutrient concentrations (focusing on TP and TN), as well as common biological parameters (focusing on Secchi transparency and chlorophyll-a) and watershed attributes (both numeric and categorical) for Texas reservoirs, providing a defensible approach from which Texas reservoirs could be grouped by watershed attributes;
3) to develop datasets of median and mean nutrient concentrations (focusing on TP) and common biological parameters (focusing on chlorophyll-a and Secchi transparency) calculated using five approaches to handling censored observations;
4) to identify nutrient threshold values associated with changes in the magnitude or variability of commonly measured biological response variables for Texas reservoirs and determine if the approach to handling censored observations in calculating measures of central tendency affected these threshold values.

2.1. RESERVOIR DATABASE DEVELOPMENT, MEDIAN CALCULATION, AND FREQUENCY DISTRIBUTIONS

Methods

Water Quality Database

Data Acquisition, Compilation and Reduction. TCEQ provided a database of water quality data collected from 1968 to 2012 from reservoirs throughout the Texas. The collected data was from 764 stations spanning 14 watersheds and was divided among three Microsoft Excel workbooks. The data described 116 reservoir characteristics and water quality parameters including nutrients, sediments, transparency, physico-chemical parameters, as well as others.

For the purposes of advanced statistical analyses conducted during this project, only data collected under specific monitoring type codes (as decided by TCEQ) and from 2000 to 2010 was used. Therefore, the database was sorted and any data collected before calendar year 2000 or after 2010 was removed. Data collected under the monitoring type code Biased Flow (BF) was also removed since data collected under this circumstance were not necessarily representative of baseline water quality conditions. The data received from TCEQ were output to a single column format within the files, so the data were reorganized into a useable format using the pivot table function in Microsoft Excel so that each parameter and the associated data were unique to an individual column. Any censored data points (i.e., those reported with a < or >) were flagged, using the numeral "1" to denote "<", "2" to denote ">", and "3" to denote a parameter calculated from multiple parameters using one or more censored values. Non-censored data points were flagged with the numeral "0." Censored values were replaced in the database with the detection limit.

Several additional parameters were calculated from the original data provided. Nitrate plus nitrite-nitrogen (NO_x-N) and total nitrogen (TN) were calculated if the necessary N species were provided by TCEQ in the original data file. In addition, diel change (i.e., 24 hour maximum minus 24 hour minimum) was calculated for dissolved oxygen (DO Flux), temperature, conductivity, pH, and turbidity. The additional parameters were added to each station worksheet using a Microsoft Excel Macro (Appendix 2.1).

Due to the volume of data provided, several parameters were removed from the median database because of lack of data and duplication of parameters, or because TCEQ indicated that the parameter could be removed from the database.

Median and Frequency Distribution Calculations. For this study, frequency distribution and, subsequently, stressor-response analyses were conducted on station medians in order to focus on broadly applicable regional and statewide trends. Because each stream and river in Texas was not equally represented in the raw water quality dataset, conducting statistical analyses on medians removes potential site-specific bias for sites that are over- or under-represented in the raw dataset. Furthermore, biological response and nutrient stressor data did not always overlap in the raw data. Conducting analyses with median values allowed comparison of long-term trends in biological and nutrient data for these stations. Median values of each parameter were calculated for each Station ID using a Microsoft Excel Macro (Appendix 2.2). Median values were calculated based on at least 10 data points, i.e. no medians were calculated if less than 10 data points were available for a given parameter at a given station. The calculated medians for each Station ID were then compiled into one database. This database was merged with the GIS and LULC data and used in advanced statistical analysis.

Frequency distributions (minimum value, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles and maximum value) for water quality parameters TP (TCEQ parameter code 00665), TN (calculated parameter code 00600C; TCEQ parameter code 00625 + 00630, 00625 + 00593 or 00625 + 00615 + 00620), NO_x-N (calculated parameter 00630C; TCEQ parameter code 00630, 00593 or 00615 + 00620), PO_4-P (TCEQ calculated parameter code 00671C; TCEQ parameter code 00671 or 70507), and sestonic chl-a (TCEQ parameter code 70953) were calculated using Microsoft Excel. For this study, a parameter combining chl-a measured spectrophotometrically (parameter code 32211; chl-a spec) and chl-a fluoro was not created due to inconsistencies between the methods (Laurie Eng, personal communication). Data were more complete and censorship was less of a concern for chl-a fluoro than for chl-a spec. Spectrophotometric chl-a data were commonly censored at a relatively high detection limit (10 µg/L). Analysis exploring the effects of censored data on chl-a spec median calculation in Texas reservoirs indicated that, when censored data exceeded 16% of the raw data for a station, chl-a medians (i.e. the 50th percentile) were increasingly overestimated when the detection limit was substituted for censored observations as the level of censoring increased because the median could not be calculated to be a value below the detection limit (see Section 2.3). This censored data effect would have been magnified further when considering very low percentiles in the frequency distribution, such as the 25th percentile. Therefore, frequency distributions for sestonic chl-a were only calculated for the fluorometric method in this study. Frequency distributions were calculated for the general streams population at multiple spatial scales including basin, level III ecoregion, basin by level III ecoregion (i.e., unique combinations of basin and level III ecoregions combined), and level IV ecoregion. Frequency distributions were also calculated for the least disturbed stations.

Geospatial Database

A geospatial database contained within a Microsoft Excel file was provided by TCEQ that identified land use and land cover data for the water quality stations located within reservoirs included in this study. The geospatial descriptors were provided for the reservoir drainage basin. Reservoir drainage basins were delineated using aggregate HUC 12 sub-watersheds, and constrained by the nearest upstream reservoir or headwater drainage boundary. Typically drainage sizes were larger than HUC-8 boundaries, but smaller than HUC-6. The descriptors included percent open water, developed-open, developed-low intensity, developed- medium intensity, developed-high intensity, barren land, deciduous forest, evergreen forest, mixed forest, shrub/scrubland, grassland/herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands. These descriptors were reduced to five categories including percent developed (i.e., open, low intensity, medium intensity, barren land), forest (i.e., deciduous, evergreen, mixed and shrubland), agriculture (i.e., grassland/herbaceous, pasture/hay, cultivated crops), developed plus agriculture (i.e., open, low intensity, medium intensity, barren land, grassland/herbaceous, pasture/hay, cultivated crops) and wetlands (i.e., woody and emergent herbaceous wetlands). Additional geospatial information for each site was provided including drainage area, slope, municipal discharges, basin ID, level III ecoregion ID, and level IV ecoregion ID.

Data Quality Assurance and Control

Data quality checks were employed frequently throughout the database reorganization and data calculation processes. The original source files were maintained in an unaltered form, and subsequent changes to each database were saved under unique file names. Data transferred from one file to the next were checked for accuracy by comparing first and last rows and the row count between files. In addition, when calculations were preformed, including manual calculations and those calculated using Microsoft Excel Macros, at least 10 percent of calculations were checked for accuracy following the secondary data quality assurance project plan (QAPP).

Results and Discussion

Frequency distributions of nutrient and sestonic chl-a medians calculated for all reservoir stations which represent general reservoir conditions were calculated at multiple spatial scales. The frequency distributions of median nutrient and chlorophyll-a concentrations from reservoirs among basins and level III ecoregions are discussed here (Tables 2.1.1 and 2.1.2). The frequency distributions of median nutrient and chl-a concentrations from basin x level III ecoregion and level IV ecoregion are presented in Appendix 2.3.

Basin

The state of Texas is divided into 23 basins which are categorized as river (65%) or coastal (35%) basin waters. River basin waters are the surface inland waters comprising the major streams, reservoirs and their tributaries, while coastal basin waters are surface inland waters that discharge or in some way interconnect with bays or the Gulf of Mexico. The 25th percentile of the median TP concentrations was less than 0.060 mg/L, a common detection limit in the database, at 60% of the Texas basins, and the 25th percentiles at these basins ranged from 0.050-0.170 mg/L. Two basins had insufficient data to calculate percentiles. Stations with 25th percentile TP concentrations exceeding 0.060 mg/L consistently had fewer than 20 medians, fewer than the 30 data points recommended by the USEPA as a minimum for analyzing frequency distributions to guide nutrient criteria development (USEPA 2000). The 25th percentile of the median PO_4-P concentration ranged from 0.009-0.053 mg/L for Texas reservoirs. The 25th percentiles of TP and PO_4-P concentrations were weakly positively correlated (r^2=0.13, p=0.12).

Less TN data was available for analysis, so frequency distributions were only calculated for 70% of the basins. Only three of the basins had more than 30 medians for calculation of frequency distributions. The 25th percentile of the median TN concentrations ranged from 0.50-1.08 mg/L. The 25th percentile of median NOx-N concentrations ranged from 0.01-0.07, but could not be calculated due to limited data for three basins. The 25th percentiles of TN and NOx-N were weakly negatively correlated (r^2=0.12, p=0.14). As with TN concentrations, the 25th percentile of sestonic chl-a data could only be calculated for 70% of the basins. Only two basins had more than 30 chl-a medians for calculating a distribution. These basins had moderate 25th percentile values, both approximately 10 µg/L. The 25th percentile of chl-a concentrations ranged from 3.00-26.1 µg/L across all basins. The 25th percentile of sestonic chl-a concentrations was 2-8x greater for reservoirs in Basin 3 than for those in any other Texas basin. Though Basin 3 did not have the lowest number of medians contributing to frequency distribution calculations for sestonic chl-a, the number of data points was low for this basin, only n=8. The 25th percentiles of the median concentrations of sestonic chl-a and TN were positively correlated (r^2=0.40, p=0.012). Surprisingly, no other correlation between median sestonic chl-a and nutrient concentration 25th percentiles was found.

Table 2.1.1. Frequency distribution of median nutrient and chlorophyll-a concentrations from reservoirs among basins in Texas, 2000-2012; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	4	0.050	--	0.058	0.060	0.060	--	0.060
2	20	0.050	0.050	0.060	0.077	0.166	0.211	1.140
3	10	0.080	0.090	0.100	0.126	0.148	0.168	0.190
4	14	0.040	0.050	0.050	0.053	0.060	0.079	0.100
5	12	0.050	0.060	0.060	0.060	0.073	0.080	0.095
6	31	0.050	0.050	0.060	0.060	0.060	0.100	0.190
8	78	0.015	0.031	0.047	0.060	0.087	0.157	0.720
10	17	0.100	0.156	0.170	0.200	0.280	0.366	0.440
12	64	0.020	0.025	0.060	0.060	0.070	0.137	0.495
14	47	0.020	0.046	0.058	0.060	0.060	0.060	0.140
18	1	--	--	--	0.050	--	--	--
19	1	--	--	--	0.060	--	--	--
21	5	0.057	--	0.070	0.140	0.166	--	0.208
23	6	0.050	--	0.050	0.055	0.060	--	0.060

Total Nitrogen (TN; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	4	0.47	--	0.50	0.57	0.65	--	0.69
2	11	0.51	0.53	0.58	0.71	0.84	2.18	8.94
3	8	1.04	1.05	1.08	1.09	1.11	1.15	1.21
4	13	0.53	0.56	0.60	0.70	0.77	0.79	1.01
5	11	0.53	0.83	0.99	1.00	1.12	1.14	1.19
6	18	0.38	0.44	0.56	0.74	1.01	1.33	3.32
8	67	0.36	0.56	0.80	0.95	1.05	1.20	6.59
10	1	--	--	--	1.16	--	--	--
12	57	0.53	0.63	0.79	1.14	1.47	1.73	2.05
14	38	0.31	0.44	0.54	0.63	0.86	1.13	2.45
18	0	--	--	--	--	--	--	--
19	1	--	--	--	0.41	--	--	--
21	3	0.49	--	--	0.99	--	--	1.08
23	6	0.46	--	0.64	1.01	1.53	--	3.11

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	4	0.04	--	0.04	0.04	0.04	--	0.05
2	14	0.04	0.04	0.04	0.05	0.05	0.10	5.70
3	8	0.04	0.04	0.04	0.05	0.05	0.05	0.05
4	13	0.02	0.04	0.04	0.05	0.05	0.05	0.05
5	9	0.04	0.04	0.04	0.04	0.05	0.08	0.15
6	31	0.04	0.04	0.05	0.05	0.07	0.11	1.94

8	67	0.00	0.01	0.01	0.04	0.08	0.21	4.47
10	1	--	--	--	0.19	--	--	--
12	45	0.01	0.04	0.04	0.05	0.07	0.14	0.51
14	40	0.02	0.02	0.03	0.05	0.09	0.18	0.32
18	1	--	--	--	0.10	--	--	--
19	1	--	--	--	0.07	--	--	--
21	5	0.02	--	0.02	0.02	0.02	--	0.03
23	6	0.04	--	0.07	0.14	0.22	--	0.44

Ortho-Phosphate (PO_4-P; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	4	0.040	--	0.040	0.040	0.045	--	0.060
2	20	0.040	0.040	0.040	0.060	0.103	0.122	0.910
3	10	0.007	0.008	0.025	0.060	0.060	0.060	0.060
4	17	0.010	0.020	0.020	0.050	0.060	0.060	0.060
5	12	0.010	0.040	0.040	0.040	0.040	0.040	0.050
6	26	0.040	0.040	0.040	0.040	0.040	0.060	0.060
8	93	0.002	0.006	0.009	0.013	0.030	0.060	0.455
10	14	0.040	0.042	0.053	0.080	0.115	0.204	0.345
12	86	0.002	0.006	0.020	0.040	0.040	0.055	0.270
14	47	0.010	0.010	0.020	0.040	0.040	0.060	0.060
18	0	--	--	--	--	--	--	--
19	1	--	--	--	0.040	--	--	--
21	4	0.010	--	0.048	0.066	0.075	--	0.083
23	6	0.040	--	0.040	0.040	0.040	--	0.050

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
1	3	5.30	--	--	5.89	--	--	19.0
2	12	3.68	7.39	8.84	13.4	24.0	61.3	82.9
3	8	16.2	18.9	26.1	29.5	34.2	45.6	71.1
4	9	5.60	5.90	7.47	16.9	20.1	22.1	28.9
5	3	25.9	--	--	44.5	--	--	49.6
6	19	3.11	4.30	10.2	13.0	33.9	39.4	52.4
8	9	4.57	8.07	10.5	15.9	17.6	29.9	30.7
10	4	6.78	--	9.08	13.1	20.7	--	33.6
12	50	3.00	3.73	10.4	16.9	23.0	29.5	69.4
14	42	0.130	0.922	5.00	7.65	11.8	16.2	53.3
18	0	--	--	--	--	--	--	--
19	1	--	--	--	3.00	--	--	--
21	4	5.35	--	12.0	15.2	16.5	--	17.1
23	5	3.00	--	3.00	3.05	21.9	--	31.8

Level III Ecoregions

Texas is divided into 11 level III ecoregions comprised of deserts (9%), tablelands (9%), timbers (9%), plateaus (9%), prairies (9%), and plains (55%). The 25th percentiles of median TP and PO4-P concentrations were calculated for all of the level III ecoregions, except for ecoregion 25-High Plains where only 2 medians were available. Of the remaining ecoregions, five had less than 30 datapoints for both parameters. Within level III ecoregions, the 25th percentile of median TP concentrations exceeded 0.060 mg/L only for ecoregion 31-Southern Texas Plains. With only six medians, this ecoregion had the second fewest datapoints. The 25th percentile of TP medians displayed minimal variability and ranged from 0.029-0.063 mg/L. The 25th percentile of the median PO4-P concentrations ranged from 0.008-0.040 mg/L.

Insufficient data was available to calculate the 25th percentile of median TN or NO_x-N concentrations for ecoregion 25-High Plains. Of the remaining ecoregions, only 2-3 had greater than 30 medians for calculating frequency distributions. The 25th percentiles of median TN concentrations ranged from 0.45-0.95 mg/L, while the 25th percentiles of median NO_x-N concentrations ranged from 0.02-0.14. Both the highest and lowest NO_x-N concentration ecoregions had the lowest number of datapoints (n=5-6). Data was also insufficient to calculate frequency distributions of median sestibuc chlorophyll-a concentrations for ecoregion 25-High Plains. Only two ecoregions had greater than 30 chl-a datapoints to calculate percentiles. The 25th percentile of the median sestonic chl-a concentrations ranged from 3.00-16.2 µg/L.

Table 2.1.2. Frequency distribution of median nutrient and chlorphyll-a concentrations from reservoirs among level III ecoregions in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	5	0.050	--	0.050	0.050	0.060	--	0.060
25-High Plains	2	0.060	--	--	0.063	--	--	0.065
26-Southwestern Tablelands	11	0.050	0.055	0.060	0.060	0.060	0.085	1.140
27-Central Great Plains	23	0.050	0.052	0.060	0.060	0.155	0.197	0.220
29-Cross Timbers	92	0.020	0.031	0.050	0.060	0.070	0.080	0.155
30-Edwards Plateau	27	0.040	0.046	0.060	0.060	0.060	0.060	0.068
31-Southern Texas Plains	6	0.057	--	0.063	0.105	0.160	--	0.208
32-Texas Blackland Prairies	30	0.015	0.020	0.029	0.050	0.068	0.188	0.270
33-East Central Texas Plains	34	0.032	0.050	0.060	0.070	0.119	0.215	0.720
35-South Central Plains	80	0.040	0.050	0.060	0.068	0.160	0.200	0.440

Total Nitrogen (TN; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	5	0.46	--	0.59	0.80	1.21	--	1.63
25-High Plains	2	0.62	--	--	1.34	--	--	2.05
26-Southwestern Tablelands	9	0.47	0.50	0.53	0.63	0.90	3.37	8.94
27-Central Great Plains	17	0.51	0.69	0.73	0.89	1.05	1.69	2.45
29-Cross Timbers	78	0.47	0.58	0.66	0.87	1.04	1.28	1.85
30-Edwards Plateau	21	0.31	0.41	0.45	0.55	0.62	0.96	1.27

31-Southern Texas Plains	4	0.49	--	0.86	1.04	1.59	--	3.11
32-Texas Blackland Prairies	22	0.36	0.50	0.72	0.94	1.32	1.60	1.76
33-East Central Texas Plains	32	0.59	0.84	0.95	1.12	1.37	1.73	6.59
35-South Central Plains	48	0.38	0.54	0.60	0.87	1.09	1.24	3.32

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	5	0.04	--	0.14	0.15	0.24	--	0.44
25-High Plains	2	0.05	--	--	0.28	--	--	0.51
26-Southwestern Tablelands	11	0.02	0.02	0.03	0.04	0.04	0.05	5.70
27-Central Great Plains	17	0.02	0.04	0.04	0.05	0.05	0.06	0.16
29-Cross Timbers	75	0.00	0.01	0.03	0.04	0.05	0.11	0.33
30-Edwards Plateau	20	0.02	0.02	0.03	0.05	0.10	0.14	0.18
31-Southern Texas Plains	6	0.02	--	0.02	0.02	0.02	--	0.05
32-Texas Blackland Prairies	21	0.00	0.04	0.05	0.08	0.11	0.24	0.32
33-East Central Texas Plains	28	0.01	0.02	0.04	0.05	0.13	0.18	4.47
35-South Central Plains	60	0.02	0.04	0.04	0.05	0.05	0.15	1.94

Ortho-Phosphate (PO_4-P; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	5	0.040	--	0.040	0.040	0.040	--	0.050
25-High Plains	2	0.040	--	--	0.050	--	--	0.060
26-Southwestern Tablelands	11	0.020	0.020	0.040	0.040	0.055	0.060	0.910
27-Central Great Plains	25	0.010	0.040	0.040	0.060	0.095	0.116	0.135
29-Cross Timbers	114	0.002	0.005	0.008	0.020	0.040	0.040	0.060
30-Edwards Plateau	26	0.010	0.010	0.020	0.038	0.040	0.040	0.040
31-Southern Texas Plains	5	0.010	--	0.040	0.060	0.072	--	0.083
32-Texas Blackland Prairies	37	0.006	0.010	0.010	0.020	0.040	0.060	0.120
33-East Central Texas Plains	35	0.008	0.010	0.010	0.040	0.055	0.060	0.455
35-South Central Plains	80	0.007	0.038	0.040	0.048	0.060	0.090	0.345

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	5	3.00	--	3.00	3.05	21.9	--	31.8
25-High Plains	2	7.75	--	--	26.6	--	--	45.5
26-Southwestern Tablelands	9	3.68	4.97	5.89	13.6	19.0	41.9	65.3
27-Central Great Plains	14	3.44	5.07	7.20	12.1	15.7	42.7	82.9
29-Cross Timbers	48	3.00	4.11	9.36	16.2	21.8	25.4	29.7
30-Edwards Plateau	25	0.130	0.484	3.00	5.45	8.30	12.4	15.7
31-Southern Texas Plains	4	5.35	--	12.0	15.2	16.5	--	17.1
32-Texas Blackland Prairies	10	0.250	1.47	3.64	9.78	17.2	31.4	38.2
33-East Central Texas Plains	10	5.97	7.21	16.2	22.2	25.9	42.2	69.4
35-South Central Plains	42	3.11	5.85	10.2	19.3	33.4	42.9	71.1

Aggregate Ecoregions

The USEPA has suggested nutrient criteria for TP, TN and chl-a for the 14 aggregate ecoregions in the United States based upon frequency distributions of found data. The USEPA recommendations for the aggregate ecoregions that lie within Texas are presented in Table 2.2.3. For TP concentrations, the EPA suggested nutrient criteria for these aggregate ecoregions ranged from 0.017-0.038 mg/L. All but one of the 25th percentile estimates for TP medians in Texas basins and level III ecoregions exceeded this range. For TN concentrations, the USEPA suggested nutrient criteria for aggregate ecoregions within Texas ranged from 0.36-0.78 mg/L. Most of the 25th percentile estimates for TN medians in Texas basins and level III ecoregions were within this range, though Basin 3 exceeded the maximum by greater than 25%. For chlorophyll-a concentrations, the USEPA suggested criteria for aggregate ecoregions within Texas ranged from 2.00-8.59 µg/L. The 25th percentile of chl-a medians in Texas basins and level III ecoregions almost always exceeded this range.

Table 2.1.3. USEPA recommended criteria for total phosophorus, total nitrogen, and chlorophyll-a for lakes and reservoirs in each of the aggregate ecoregions that lie partially within Texas.

Aggregate Ecoregion	Total Phosphorus (mg/L)	Total Nitrogen (mg/L)	Chlorophyll-a (µg/L)
Aggregate Ecoregion II	0.017	0.40	3.40
Aggregate Ecoregion III	0.02	0.44	2.00
Aggregate Ecoregion IV	0.033	0.56	2.30
Aggregate Ecoregion V	0.038	0.78	8.59
Aggregate Ecoregion IX	0.020	0.36	4.93

The development of frequency distributions from median parameter concentrations is an important first step in the development of nutrient criteria, and the 25th and 75th percentile method recommended by the USEPA (2000) should be used as a guide when setting criteria for specific geospatial regions. The frequency distribution is also a good method to estimate the number of sites within a spatial scale (e.g., basins, ecoregions) that could exceed the developed criteria. However, this study, as well as others (Ice et al. 2003; Binkley 2004; Longing and Haggard 2010), have shown that the 25th percentile frequency distribution can vary from one basin or ecoregion to another and different spatial scales. These studies have shown that 25th percentiles based on regional data often significantly differ from that developed for the aggregate ecoregions. The frequency distribution method should only be one of many tools used to support the development of numeric nutrient criteria. The Science Advisory Board has advised the USEPA that the stressor-response approach is a legitimate, scientifically based method for developing nutrient criteria when correctly applied, and is this approach is the focus of subsequent analyses in this chapter.

2.2 GROUPING RESERVOIR STATIONS WITH SIMILAR NUTRIENT AND BIOLOGICAL CONDITIONS BY THRESHOLDS IN GEOSPATIAL VARIABLES

Methods

We conducted Classification and Regression Tree (CART) analyses on the median database for Texas reservoirs to group reservoir stations by watershed attributes by similar nutrient and/or biological conditions. The focus (dependent) nutrient variables of these analyses were median TP, TN, and spectrophotometric and fluorometric chlorophyll-a concentrations (TCEQ parameter codes 00665, 00600C, 32211, and 70953, respectively. The watershed attributes considered as predictor (independent) variables were divided into 3 categories: 1) land use/land cover (LULC), 2) permitted municipal waste water treatment plant flow, and 3) regions. LULC was further divided into categories, including percent developed, percent agriculture, percent developed + agriculture, percent forested, and percent wetland. Each of these LULC categories combined several TCEQ LULC codes, as described as part of the development of the geospatial database in Section 2.1. Waste water treatment plant discharge was used as a predictor variable both unweighted (mgd) and weighted by watershed area (mgd/km^2). Because background discharge volumes for Texas streams and rivers were not available, we divided permitted municipal WWTP flow by the catchment area to approximate the proportion of in-stream flow supplied by municipal discharge. Regions considered in the analysis were level III ecoregion, basin, and a combination of basin by level III ecoregion.

CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). For the models with the greatest explanatory power, CART analyses were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value (see Section 2.1). A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 2.4 the statistical code and raw output generated for each CART and nCPA analysis conducted for this study has been compiled.

Results and Discussion

Land use/Land cover

LULC categories were weak predictors for median TP and TN concentrations in Texas reservoirs (max r^2 = 0.13) but somewhat stronger predictors for biological variables (max r^2=0.20). The %Developed LULC was the strongest predictor among LULC categories for TP (Fig. 2.2.1A), indicating that TP concentrations were approximately double on average above a threshold of 13% developed land in the watershed. For TN, %Agriculture was the strongest predictor among LULC categories (Fig. 2.2.1B). Above a threshold of 57% agricultural land, TN concentrations increased by approximately 35%. For all biological variables considered, CART analyses indicated thresholds in %Agriculture, ranging from 18% for Secchi transparency to 22-38% for chlorophyll-a (Figs. 2.2.2A-C). These models had the greatest explanatory power of all geospatial models using LULC categories as explanatory variables (r^2 ~ 0.20).

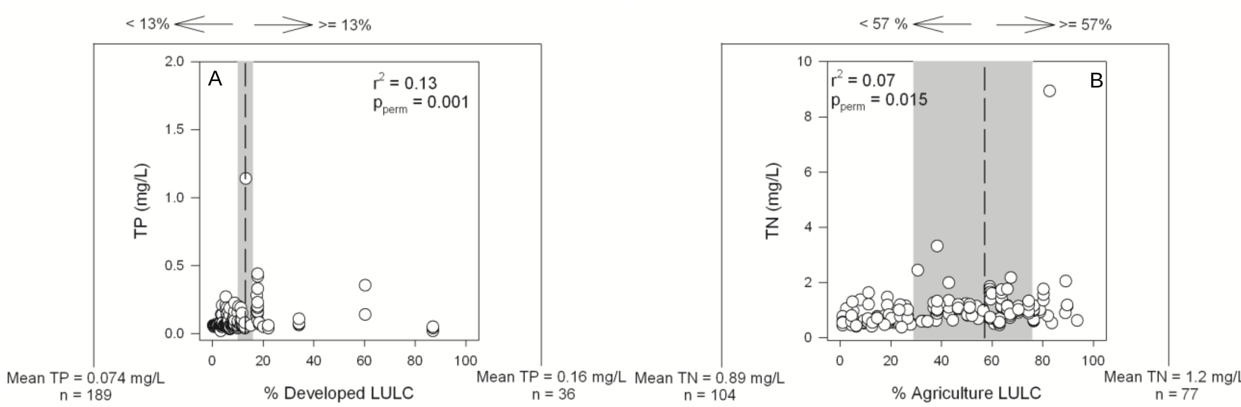


Figure 2.2.1. CART model of median total phosphorus (TP) (A) and total nitrogen (TN) (B) concentrations vs. LULC categories (%Developed, %Agriculture, %Developed + Agriculture, %Forest, and %Wetland) for Texas reservoirs.

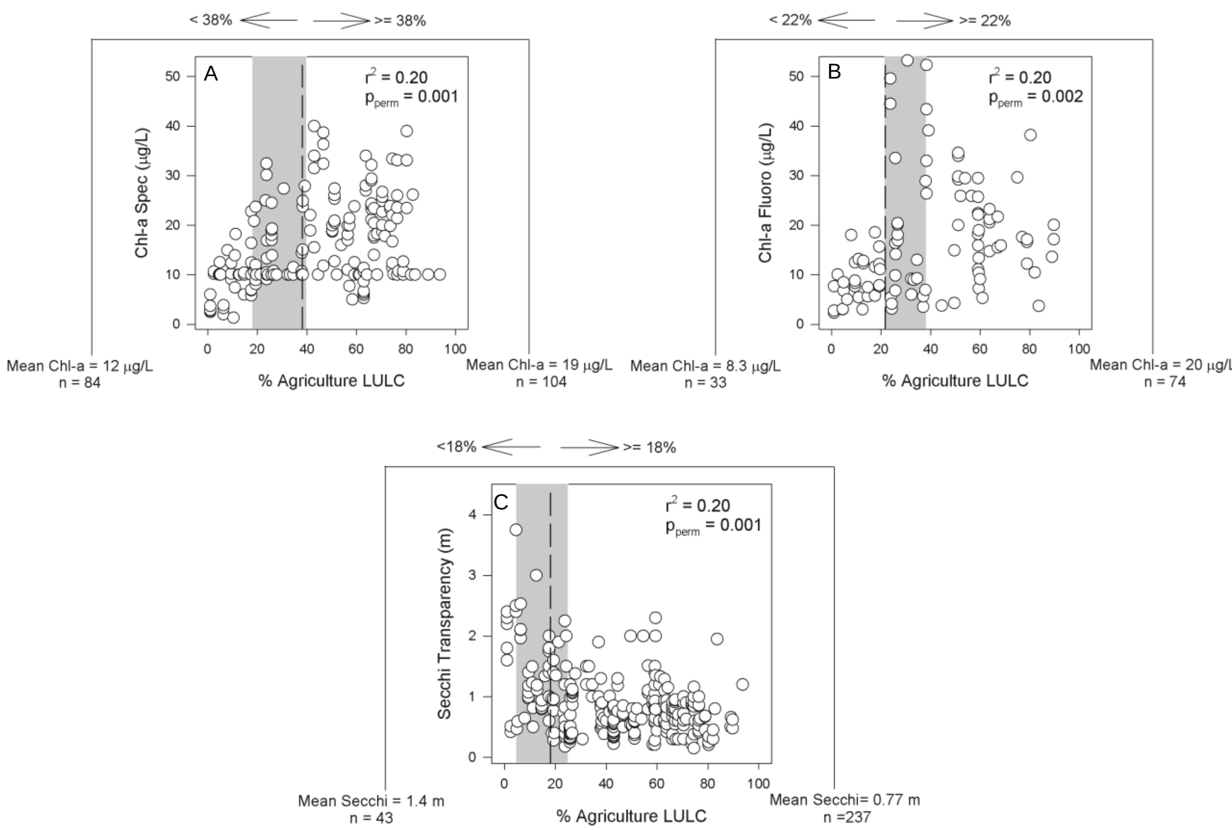


Figure 2.2.2. CART model of sestonic spectrophotometric (A) and fluorometric (B) chlorophyll-a (chl-a spec and fluoro, respectively) concentrations and Secchi transparency (C) vs. LULC categories (%Developed, %Agriculture, %Developed + Agriculture, %Forest, and %Wetland) for Texas reservoirs.

Permitted Municipal Wastewater Treatment Plant Flow

Geospatial models based on WWTP flow performed similarly both for unweighted and weighted for watershed area (Figs. 2.2.6-2.2.11; r^2 = 0.08-0.16). In general, WWTP flow was a weak predictor for nutrient concentrations and biological parameters in Texas reservoirs. For TP, CART analyses identified thresholds of 150 mgd and 0.028 mgd/km^2 (Fig. 2.2.3A-B). Both municipal discharge resulted in more than doubling of TP concentrations on average when the threshold was exceeded. The same unweighted WWTP flow threshold was identified for Secchi transparency (Fig. 2.2.3C). Exceeding 150 mgd, result in a greater than 50% reduction reservoir transparency on average. The WWTP flow thresholds identified in these three analyses represented relatively high loading rates that were exceeded by only a small number of stations (n = 14 – 33).

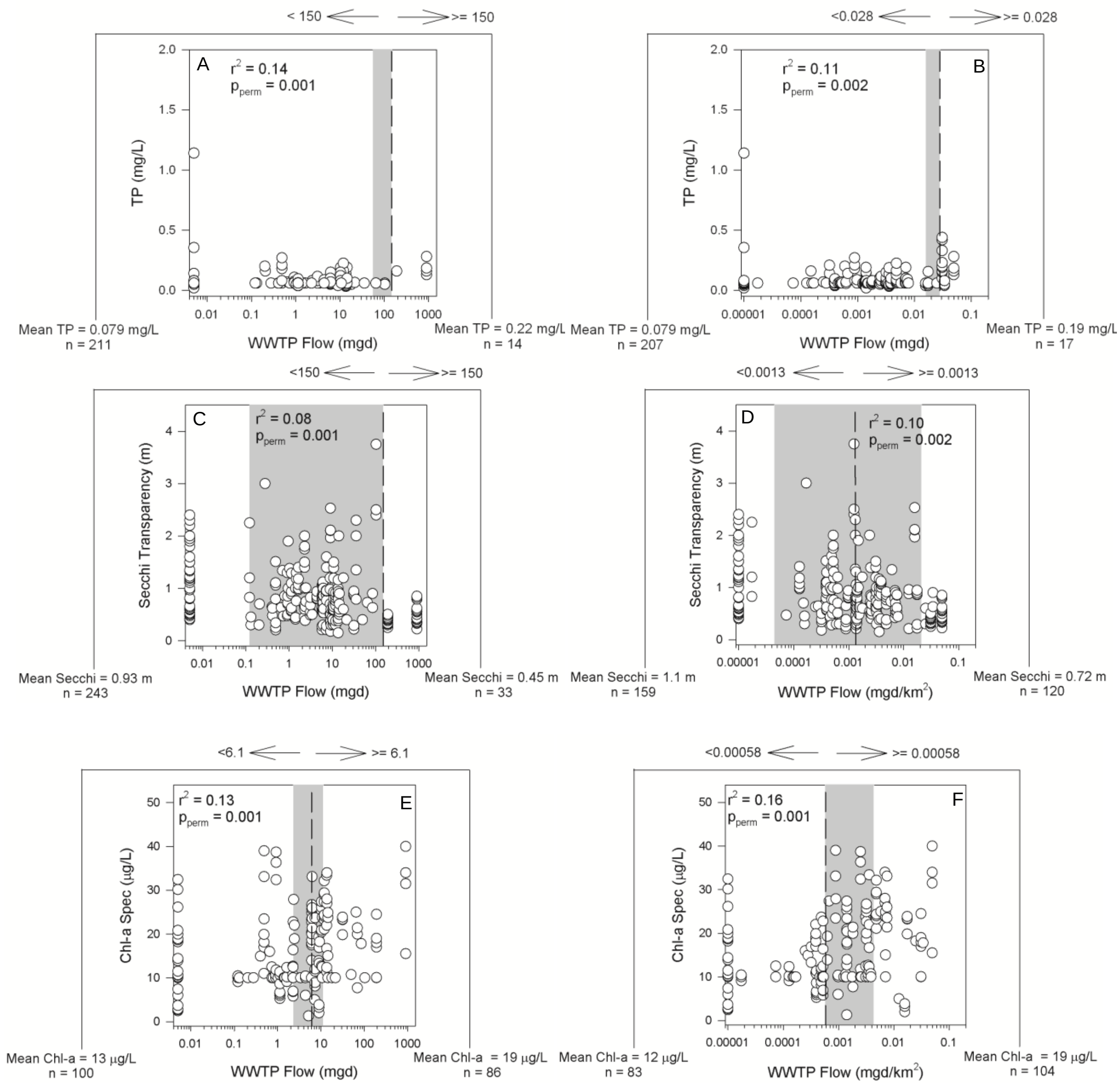


Figure 2.2.3. CART model of TP (A-B), Secchi transparency (C-D), and sestonic spectrophotometric chlorophyll-a (chl-a spec) (E-F) vs. permitted municipal WWTP flow both unweighted and weighted by watershed area for Texas reservoirs.

In contrast, the area-weighted permitted municipal WWTP flow thresholds identified for Secchi transparency and spectrophometric chlorophyll-a, as well as the unweighted flow threshold identified for Secchi transparency, were lower than those identified for TP concentration (Figs. 2.2.3D-F). Average response variable conditions above and below these loading rates diverged less than TP concentrations above and below 150 mgd or 0.028 mgd/km^2, but larger populations of “high” WWTP flow sites were generated with these analyses. These findings may indicate that, for Texas reservoirs, the effects of low-level permitted municipal WWTP flow are primarily evident as enhanced productivity rather than consistently elevated nutrient concentrations, which were only observed at high loading rates. No statistically significant splits were identified for TN or fluorometric chlorophyll-a concentrations using either unweighted or weighted discharge.

Regions

Grouping schemes based on regions had the greatest explanatory power among geospatial categories for both nutrient and biological parameters (r^2=0.14-0.31). For all variables, basin by ecoregion III was the strongest predictor. As with permitted municipal WWTP flow, however, CART separated only a few very "high" nutrient stations (n= 15 – 22), and only four basin by ecoregion areas from the rest of the reservoirs stations.

Table 2.2.1. CART groupings of Basin-Ecoregion III by "low" and "high" median TP and TN concentrations for Texas reservoirs.

Median TP r^2 = 0.27		Median TN r^2 = 0.14	
"Low", n = 204 Mean TP = 0.074 mg/L Min = 0.02 mg/L Max = 0.36 mg/L	"High", n = 22 Mean TP = 0.24 mg/L Min = 0.055 mg/L Max = 1.1 mg/L	"Low", n = 167 Mean TN = 0.80 mg/L Min = 0.38 mg/L Max = 3.3 mg/L	"High", n = 15 Mean TN = 1.9 mg/L Min = 0.53 mg/L Max = 8.94 mg/L
1-26	2-26	1-26	2-26
2-25	8-35	2-25	12-25
2-27	10-35	2-27	12-32
2-29	12-32	2-29	12-33
2-33		2-33	
3-35		3-35	
4-33		4-33	
4-35		4-35	
5-33		5-33	
5-35		5-35	
6-35		6-35	
8-29		8-29	
8-32		8-32	
8-33		8-33	
12-25		8-35	
12-26		10-35	
12-27		12-26	
12-29		12-27	
12-30		12-29	
12-33		12-30	
14-26		14-26	
14-27		14-27	
14-29		14-29	
14-30		14-30	
14-32		14-32	
18-30		18-30	
21-31		21-31	
23-31		23-31	
23-24		23-24	

For chlorophyll-a measured using both methods and for Secchi transparency, a greater number of "high" stations were identified and basin by ecoregion areas were identified relative to TN and TP. The "low" and "high" basin by ecoregion groupings are summarized in Tables 2.2.1-2.2.2.

Table 2.2.2. CART groupings of Basin-Ecoregion III by "low" and "high" median chlorophyll-a measured spectrophotometrically and fluorometrically (chl-a spec and fluoro) and Secchi transparency for Texas reservoirs.

Median Chl-a Spec $r^2 = 0.31$		Median Chl-a Fluoro $r^2 = 0.22$		Median Secchi $r^2 = 0.27$	
"Low", n= 102 Mean = 11 µg/L Min = 1.3 µg/L Max = 27 µg/L	"High", n= 87 Mean = 21 µg/L Min = 5 µg/L Max = 40 µg/L	"Low" n= 89 Mean = 14 µg/L Min = 2.3 µg/L Max = 53 µg/L	"High" n= 19 Mean = 29 µg/L Min = 3.7 µg/L Max = 65 µg/L	"Low" = 244 Mean = 0.77 m Min = 0.15 m Max = 2.3 m	"High" n= 37 Mean = 1.6 Min = 0.28 m Max = 3.75 m
1-26	2-29	1-26	2-26	1-26	2-25
2-25	3-35	4-33	2-33	2-27	2-26
2-26	5-33	4-35	3-35	2-33	2-29
2-27	5-35	6-35	5-33	3-35	4-33
2-33	8-29	8-32	5-35	4-35	12-30
4-33	8-32	8-35	8-29	5-33	14-30
4-35	8-33	10-35	12-32	6-35	14-32
6-35	8-35	12-26	12-33	8-29	18-30
12-26	10-35	12-29		8-32	23-24
12-29	12-32	12-30		8-33	
14-27	12-33	14-27		8-35	
14-29	14-26	14-30		10-35	
14-30		14-32		12-25	
14-32		18-30		12-26	
18-30		21-31		12-29	
21-31		23-24		12-32	
23-24				12-33	
23-31				14-26	
				14-27	
				14-29	
				21-31	
				24-31	

Geospatial Analysis Summary

A summary of the CART models for grouping nutrients and biological conditions in Texas reservoirs across the three broad geospatial categories is available in Table 2.2.3. In general, geospatial grouping schemes were more effective for biological parameters than for nutrients, especially TN concentration. This finding may reflect the fact that lakes and reservoirs have extended hydraulic residence times, causing nutrient concentrations in these systems to strongly reflect internal transformation, storage, release and removal processes, in addition to external sources. Therefore, variables indicative of system-level biological response may be appropriate for geospatial grouping schemes in lakes and reservoirs.

Table 2.2.3. Summary of CART geospatial models for Texas reservoirs with the greatest statistical power in each geospatial category (LULC, WWTP Flow, and Region).

Parameter	Geospatial Category	Predictor	Threshold	Model r^2
TP	LULC	% Developed	13%	0.13
	WWTP Flow	Unweighted	150 mgd	0.14
	Region	Basin-Ecoregion III	Groups	0.27
TN	LULC	% Agriculture	57%	0.07
	WWTP Flow	NS	NS	NS
	Region	Basin-Ecoregion III	Groups	0.14
Secchi	LULC	% Agriculture	18%	0.20
	WWTP Flow	Weighted	0.0013 mgd/km^2	0.10
	Region	Basin-Ecoregion III	Groups	0.27
Chl-a Spec	LULC	% Agriculture	38%	0.20
	WWTP Flow	Weighted	0.00058 mgd/km^2	0.16
	Region	Basin-Ecoregion III	Groups	0.31
Chl-a Fluoro	LULC	% Agriculture	22%	0.20
	WWTP Flow	NS	NS	NS
	Region	Basin-Ecoregion III	Groups	0.22

2.3. DEVELOPING RESERVOIR DATASETS WITH VARIABLE CORRECTION FOR CENSORING USING MULTIPLE APPROACHES TO ESTIMATE MEASURES OF CENTRAL TENDENCY IN CENSORED DATASETS

Methods

Calculating Measures of Central Tendency using Substitution, Deletion, and Statistically-based Methods for Handling Censored Observations

Common approaches for handling censored observations include deletion or substitution with either zero, the detection limit, or half the detection limit. These approaches are not statistically rigorous and can obscure existing patterns or introduce patterns to datasets that do not reflect real-world conditions. Though less commonly employed, statistically rigorous methods for analyzing censored data do exist. Methods for calculating summary statistics, such as means, medians, standard deviations, and percentiles are well-developed (Helsel 2012). These methods extract known information, or the frequency at which censored observations occur in the dataset relative to uncensored observations. These methods can be divided into two categories: 1) parametric, or methods that require assumptions about data distribution, such as normal or lognormal, and 2) non-parametric, or methods that do not require assumptions about data distribution. Maximum likelihood estimation (MLE) and regression order statistics (ROS) analyses are both parametric methods, while the Kaplan-Meier (KM) analysis is non-parametric. As a non-parametric method, KM is generally preferred to MLE or ROS, but the utility of each method varies according to number of total observations and percentage of censored observations in a dataset (Table 2.3.1; adapted from Helsel 2012). These analyses do not reliably estimate summary statistics for datasets when >80% of observations are censored.

Table 2.3.1. Summary of the conditions under which each method for calculating summary statistics in datasets with censored observations is preferred. Adapted from Helsel (2012).

	Amount of Available Data	
Percent Censored	<50 Observations	>50 Observations
< 50% censored	Kaplan-Meier	Kaplan-Meier
50-80% censored	Regression order statistics	Maximum likelihood estimate
>80% censored	Not recommended	Not recommended

Medians, means, and standard deviations were calculated for each TP (parameter code 00665), TN (parameter code 00600C), chl-a measured spectrophotometrically (parameter code 32211), and Secchi depth (parameter code 00078C) for stations with ≥ 12 observations per parameter collected between January 1, 2000 and December 31, 2010. Spectrophotometric chl-a was chosen for censored data and subsequent stressor-response analyses for reservoirs because censoring was so prevalent in the raw data relative to fluorometric chl-a. Prior to calculation of measures of central tendency, 3 subsets of the raw data were generated: 1) the detection limit was substituted for all censored observations, 2) half the detection limit was substituted for all censored observations, and 3) censored observations were deleted. In subsequent calculations we refer to these medians/means as $Med/Mean/Std_{subDL}$, $Med/Mean/Std_{sub1/2DL,}$ and $Med/Mean/Std_{Del}$. These median and mean estimates for the 4 focus parameters comprise Datasets 1-3, respectively. Because TP and chl-a were identified as having a high percentage of censored observations, two additional estimates of median and mean chl-a and TP concentrations were generated using different approaches to handling censored data. For Secchi depth and TN, $Med/Mean/Std_{subDL}$ estimates from Dataset 1 appear in Datasets 4-5.

To develop Dataset 4, medians and means were calculated using statistical methods that incorporate known information about censored observations, as outlined in Table 2.3.1. All analyses were carried out using the NADA library in 2.9.1 (http://www.r-project.org/). For stations with less than 50% censored data, medians and means were estimated through non-parametric KM analysis using the "cenfit" function. For stations with 50-80% censored data, ROS analysis was employed to estimate medians for stations with <50 observations, while MLE analysis was employed for stations with >50 observations using the "cenros" and "cenmle" functions, respectively. In subsequent calculations, we refer to these medians/means as $Med/Mean_{cen0\text{-}80\%}$. The $Med/Mean_{cen0\text{-}80\%}$ estimates from these analyses comprise Dataset 4. No estimates of medians or means from stations with >80% censoring were included in Dataset 4.

Determining the Effects of Censoring on Calculation of Measures of Central Tendency

In order to determine what percentage of censored observations potentially resulted in changes in measures of central tendency, the percent difference between $Med/Mean/Std_{subDL}$ and $Med/Mean/Std_{cen0\text{-}80\%}$ for TP and chl-a from each qualifying station was calculated:

$$\% \, Difference = \frac{Med/Mean/Std_{subDL} - Med/Mean/Std_{cen0-80\%}}{Med/Mean/Std_{subDL}} \quad \text{(Eq. 2.1)}$$

For each parameter for each qualifier station, the percent difference between the two Median/Mean/Std_{subDL} and Median/Mean/$Std_{cen0-80\%}$ estimates was compared to the percentage of censored data at that station. Thresholds in the percentage of censoring that resulted in consistent, measurable change in medians and means due to censoring were estimated using CART and non-parametric changepoint analysis for TP and chl-a medians and means in R 2.9.1. CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were crossvalidated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1. CART analyses were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model has a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of twelve observations used in calculating the median value. We verified CART results for all primary splits in regression trees using non-parametric changepoint analysis in R 2.9.1 (Qian et al. 2003, King and Richardson 2003). A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 2.5 the statistical code and raw output generated for each CART and nCPA analysis conducted for this study has been compiled.

Estimating Measures of Central Tendency for Stations with >80% Censoring Using Statistical Methods

A linear regression analysis was also conducted in SigmaPlot 12.3 on the data from stations with percent censored data exceeding the identified thresholds in order to develop models to predict medians and means for stations with >80% censoring. These estimates are referred to as $Med/Mean_{cen0\text{-}100\%}$ hereafter. Regression models predicted a percent change in $Med/Mean_{subDL}$ based on the percentage of censored data at a station. This percent change is referred to as $\%Difference_{model}$ hereafter. The $\%Difference_{model}$ was therefore calculated using the following equation:

$$\%Difference_{model} = m \times \%Censored + b \qquad \text{(Eq. 2.2)}$$

Where *m* and *b* are regression parameters that are unique values differing between chl-a and TP, and between medians and means. $Med/Mean_{corr}$ was then calculated using the following equation:

$$Med/Mean/Std_{cen0-100\%} = Med/Mean/Std_{subDL} - (Med/Mean/Std_{subDL} \times \%Difference_{model}) \text{(Eq. 2.3)}$$

$Med/Mean_{cen0\text{-}100\%}$ estimates for stations with >80% censoring were added to $Med/Mean_{cen0\text{-}80\%}$ estimates for stations with 0-80% censoring from Dataset 4 to form Dataset 5.

Summary of the Five Datasets with Variable Correction for Censoring in TP and Chl-a

Dataset 1 – This dataset contains medians and means for all stations with n ≥ 12 observations. Medians and means ($Med/Mean_{subDL}$) in this dataset were calculated after substituting the detection limit for censored observations in the raw data. Substitution with the detection limit is a common approach to handling censored data, but is not considered statistically rigorous.

Dataset 2 – This dataset contains medians and means for all stations with n ≥ 12 observations. Medians and means ($Med/Mean_{sub1/2DL}$) in this dataset were calculated after substituting half the detection limit for censored observations in the raw data. Substitution with half the detection limit is a common approach to handling censored data, but is not considered statistically rigorous.

Dataset 3 – This dataset contains medians and means for all stations with n ≥ 12 observations. Medians and means ($Med/Mean_{Del}$) in this dataset were calculated after deleting censored observations in the raw data. Deletion is a common approach to handling censored data, but is not considered statistically rigorous.

Dataset 4 – This dataset contains medians and means ($Med/Mean_{cen}$) generated using statistical methods that consider known information about censored observations, such as frequency of occurrence relative to uncensored observations, in calculating measures of central tendency. Medians and means were only included for stations with 0-80% censoring. The statistical methods used to estimate $Med/Mean_{cen}$ are peer-reviewed, published approaches to analyzing censored datasets.

Dataset 5 – This dataset contains all medians and means (Med/Mean$_{cen0\text{-}80\%}$) from Dataset 4, as well as estimates of medians and means (Med/Mean$_{cen0\text{-}100\%}$) for stations with >80% censoring. Med/Mean$_{cen0\text{-}100\%}$ estimates were calculated using the regression models shown in Fig. 2.3.1 and outlined in Eqs. 2.2-2.3. These estimates required extrapolation of the linear regression models outside of the model data range, and this method has not been peer-reviewed or published. Therefore, these estimates may not be defensible from a regulatory standpoint. However, analyses using medians and means from Dataset 5 provide important information on the potential effects of not including censor-corrected estimates of median and mean chl-a and/or TP for stations with >80% censoring.

Results and Discussion

The changepoint analysis of the percent difference in TP and chl-a medians, means, and standard deviations between Dataset 1 and Dataset 4 vs. the percentage of censored observations for a station indicated that censored data effects are only observable at 16-49% censored data or greater, depending upon the parameter and measure of central tendency (Figs. 2.3.1A-F). In general, means were less sensitive to effects of censored data, showing a difference between Datasets 1 and 3 at 45-49% (Figs. 2.3.1B & E) censored data vs. 16-38% for medians (Figs. 2.3.1A & D). Metrics of central tendency for TP also appeared to be less sensitive to the effects of censoring than for chl-a (threshold = 36-49% for TP vs. 16-36% for chl-a). For both TP and chl-a, percent difference in medians and means between Datasets 1 and 4 exhibited a linear response ($p<0.0001$) relative to the percentage of censored data above the censoring thresholds identified in changepoint analyses. Censoring effects on standard deviations (Figs. 2.3.1C & F) exhibited similar thresholds in the level of censoring to means, but the effect of highly censored datasets on standard deviations differed from both means and medians. Above threshold levels in the percent censored data, the difference in standard deviations between methods increased, but the change could be either positive or negative, whereas the change was typically only positive for means and medians above censoring thresholds.

The results of these analyses clearly indicated that the approach to handling censored observations affects the value of measures of central tendency above threshold levels of the percentage of censored data. At the highest (up to 80%) levels of censoring in a data, the difference between Dataset 1 and 4 medians and means was as high as 80% and 50%, respectively. If we assume that approximately 40% censored data denotes a threshold that above which, on average, censored observations affect measures of central tendency, approximately 50% of station TP medians/means and 25% of station chl-a spec medians/means would be overestimated by substituting the detection limit for censored observations prior to calculating measures of central tendency. This finding potentially has important implications for statistical analyses used in water quality criteria development that use measures of central tendency as input data.

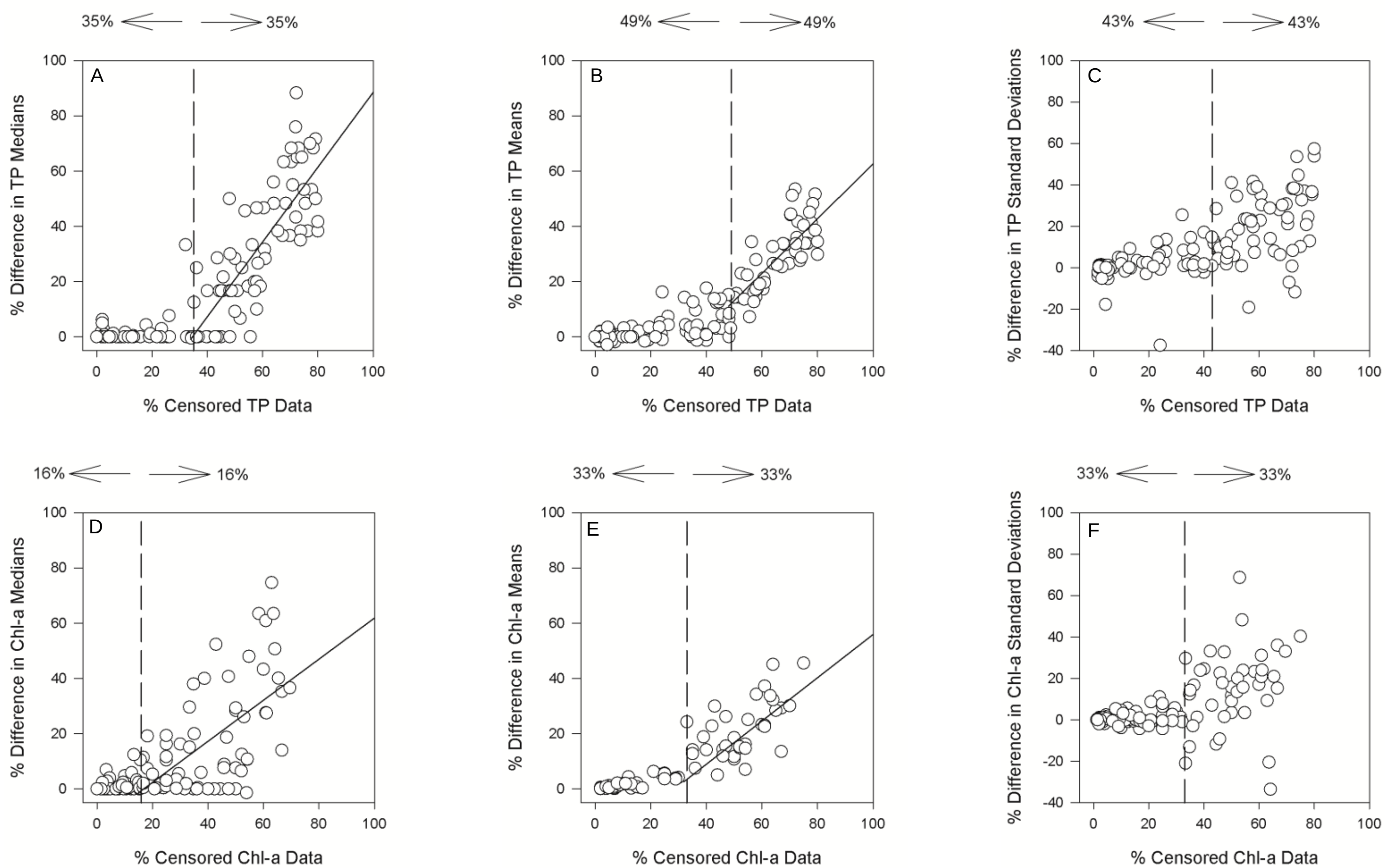


Figure 2.3.1. Results of analyses of the percent difference between reservoir station TP (A-C) and chl-a spec (D-F) medians, means, and standard deviations for each station in Datasets 1 and 4 vs. the percentage of censored observations for that station. The dashed line indicates the 5% confidence estimate for the lowest statistically significant threshold in percent censored data ($p < 0.05$) and was considered to be the level of censoring at which censored observations affected measures of central tendency. Only data falling to the right of the dashed line were included in subsequent regression analysis. The solid line indicates the regression used to estimate medians or means for Dataset 5 for stations with >80% censored observations. Changes to standard deviations above censoring thresholds were not linear ($p>0.05$).

Table 2.3.2. The frequency distribution 25th percentile estimates for all Texas reservoirs station medians estimated for total phosphorus (TP) and chlorophyll-a measured spectrophotometrically (chl-a spec) using the five approaches to handling censored data outlined in this section.

Parameter	USEPA (2000)	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
TP (mg/L)	0.07-0.038	0.06	0.03	0.068	0.041	0.019
Chl-a spec (µg/L)	2.0-8.6	10	5.0	12	8.0	6.9

For example, if the approach to handling censored data affects median calculation (i.e. the data 50th percentile), then it is possible that censoring in the dataset would also have an impact on frequency distribution percentiles calculated using these medians. An illustration of how the 25th percentile estimates of median TP and chl-a spec concentrations in Texas reservoirs varies between the 5 datasets is provided in Table 2.3.2. As with station medians, 25th percentile estimates of median TP and chl-a spec decreased in value in the datasets where statistical methods for estimating summary statistics in censored datasets were used (Datasets 4-5). For the datasets where a substitution approach was applied (substitution with the detection limit and half the detection limit in Datasets 1 and 2, respectively), the 25th percentile estimates are equal to the common detection limit or half that value for both TP and chl-a spec. Interestingly, the highest estimates of 25th percentiles occurred in Dataset 3 for both parameters. This may have occurred because insufficient data was present after deleting censored observations causing most of the stations with the lowest TP and chl-a concentrations to no longer be included in the analysis. It is also interesting to note that estimates of 25th percentiles for Dataset 2 were lower in value than estimates for Dataset 4 where corrections for censoring had been applied using statistical methods for stations with up to 80% censoring. We can conclude therefore, not only does substituting the detection limit for censored values in the raw data lead to overestimation of station medians and lower percentiles of frequency distributions using those medians, but also that, conversely, substituting half the detection limit leads to underestimation of these same values. Interestingly, all 25th percentile estimates from this study, including those that were most sensitive to censored data, fall in the upper half of the range of 25th percentiles recommended as TP and chl-a criteria by the USEPA.

2.4 STRESSOR-RESPONSE ANALYSIS ON THE RESERVOIR DATASETS WITH VARIABLE CORRECTION FOR CENSORING

Methods

We conducted classification and regression tree (CART) analyses on the 5 Texas reservoirs median and mean datasets with variable correction for censoring described in Section 2.3. Chl-a response to TP was investigated using medians and means compiled in each of the 5 datasets, i.e. 10 analyses. Median and mean Secchi depth response to TP was investigated using TP medians and means from each of the 5 datasets, i.e. 10 analyses. The chl-a response to TN was investigated using chl-a medians and means from each of the 5 datasets, i.e. 10 analyses. Finally, Secchi depth response to TN concentration was investigated using medians and means only from the datasets generated from raw data subsets (Datasets 1-3), i.e. 6 analyses, because no additional correction for censoring was carried out for these parameters.

CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were crossvalidated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). http://www.r-project.org/). For the models with the greatest explanatory power, CART analyses were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model has a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of twelve observations used in calculating the median value. We verified CART results for all primary splits in regression trees using non-parametric changepoint analysis in R 2.9.1 (Qian et al. 2003, King and Richardson 2003). A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 2.6 the statistical code and raw output generated for each CART and nCPA analysis conducted for this study has been compiled.

Results

The relationship between nutrient concentrations (particularly TP), chl-a concentrations, and Secchi transparency are well-established for lakes and reservoirs (Carlson 1977). The findings of this study were in-line with established ecological theory, indicating that Secchi transparency decreased and chl-a concentrations increased with increasing nutrient concentrations in all 5 of the Texas reservoir datasets with variable correction for censoring. While these relationships often exhibit linearity in the log scale, in this study we were able to identify numeric nutrient threshold concentrations for both TP and TN relative to chl-a and Secchi transparency.

Thresholds in median TP concentrations

A summary of TP thresholds in chl-a and Secchi depth response is provided in Table 2.4.1. Changepoint analyses of chl-a vs. TP medians and means from Datasets 1-5 indicated TP thresholds ranging from 0.039 - 0.63 mg/L (Figs. 2.4.1A-E and 2.4.2A-E). Changepoint analyses of Secchi depth vs. TP indicated a similar range of 0.025 – 0.065 mg/L (Figs. 2.4.3A-E and 2.4.4A-E). These values correspond with mid-range to upper mid-range values on Carlson’s (1977) trophic state scale for lakes and reservoirs. With only a few exceptions, these thresholds were 2-3x greater the USEPA recommended criteria for TP concentration for lakes and reservoirs in aggregate ecoregions that lie within Texas (0.07-0.38). These thresholds also exceeded many of the reference lake and reservoir conditions estimated using probability-based sampling data from the National Lakes Survey (NLA) (Herlihy et al. 2013). However, NLA reference 75th percentiles and stressor model estimated 75th percentiles for natural lakes in aggregate ecoregions IV-Grass Plains and lakes and reservoirs in V-Cultivated Great Plains were in range with or exceeded the TP thresholds estimated here.

In general, CART models based on TP concentrations had good predictive power for both chl-a (r^2=0.22-0.35) and Secchi transparency (0.19-0.44), regardless of the measure of central tendency or approach to handling censored data. In models using Datasets 1-2, where single values recurred frequently (i.e. the detection limit or half the detection limit), model predictive power may have been artificially inflated.

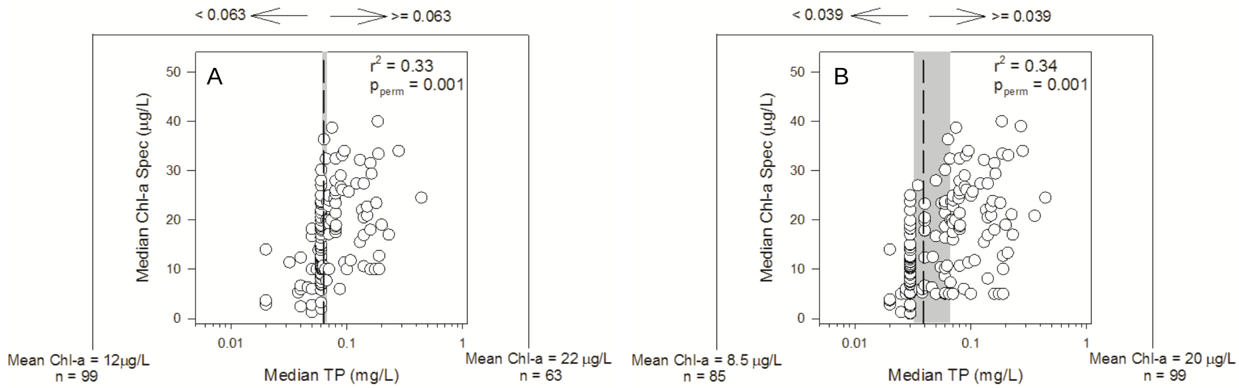
One of the most notable differences among thresholds identified in these datasets were between Dataset 1 and 2 medians. For both chl-a and Secchi depth, the value of the TP threshold was reduced by nearly half simply by substituting a lower value for censored observations. It is also important to note that, especially for medians, TP thresholds identified in changepoint analysis on Dataset 4 were higher than thresholds identified in Dataset 2. This finding suggests that substituting half the detection limit for censored observations results in underestimation of metrics of central tendency relative to estimates that can be made using statistical techniques for analyzing censored datasets that extract known information based on distribution or rank of values in the dataset. This underestimation also subsequently affected the outcome of analyses that used these measures and resulted in potential underestimation of nutrient thresholds. In general, differences in TP thresholds among datasets were less pronounced when means were the measure of central tendency used in the analysis.

Table 2.4.1. Summary of TP thresholds identified in the 5 reservoir median/mean datasets.

Dataset	Primary TP threshold (mg/L)			
	Median Chl-a	Median Secchi	Mean Chl-a	Mean Secchi
1) SubDL	0.063	0.063	0.063	0.063
2) Sub1/2DL	0.039	0.039	0.053	0.045
3) Deletion	0.060	0.039	0.060	0.065
4) Cen0-80%	0.063	0.049	0.054	0.048
5) Cen0-100%	0.049	0.025	0.054	0.042

In some instances, reductions in the TP threshold were seen when analyses used Dataset 4 medians and means. This was not the case for analyses of median chl-a vs. TP, for which the Dataset 1 and 4 TP thresholds were identical, but for median and mean Secchi depth vs. TP, an approximately 22% reduction in the TP thresholds was seen from Dataset 1 to 4. This reduction was minor, however, and may not be detectable in practice. In general, good agreement was seen between substituting with the detection limit and peer-reviewed statistical techniques for analyzing censored datasets in the Texas reservoir data. These findings suggest that substituting the detection limit for censored data results in reasonable approximation of environmental dynamics in Texas reservoirs. It is important to note that this may not be the case for other types of analyses or for other data populations.

However, study results also suggested that TP thresholds could actually be lower than the common detection limits in the Texas reservoir water quality database or as indicated by analyses on summary statistics that were calculated using established techniques for handling censored datasets (Datasets 1-4). These thresholds may only be possible to determine if the database includes estimates of median and mean TP concentrations for sites that hypothetically have the lowest TP concentrations (i.e. sites with >80% censoring). In this study, exploratory estimates were made for these stations to investigate this question. Differences in the TP thresholds, therefore, between Datasets 4 and 5 would suggest that not including censor-corrected median and mean estimates for these stations results in potential misinterpretation of stressor-response relationships. For chl-a medians, a reduction in the TP threshold was seen between analyses with Dataset 4 and 5 medians (but not means). This reduction was minor, approximately 22%. The largest difference in TP thresholds between Datasets 4 and 5 occurred in analysis of median Secchi transparency vs. TP (0.063 mg/L reduced to 0.025 mg/L). While the specific value of this threshold would not be suitable for setting criteria due to the exploratory nature of the procedures for estimating medians and means for stations with a high percentage of censored observations, this finding does suggest that including reliable estimates of TP concentrations for low-nutrient stations with a high percentage of censored data may be necessary to evaluate stressor-response relationships in Texas reservoirs effectively. To determine TP thresholds with greater certainty would require beginning to analyze samples to lower detection limits. The analyses in Section 2.3 suggest stations with >40% censoring should be target sites for more sensitive analyses because summary statistics for stations with lower levels of censoring were minimally affected by censored data.



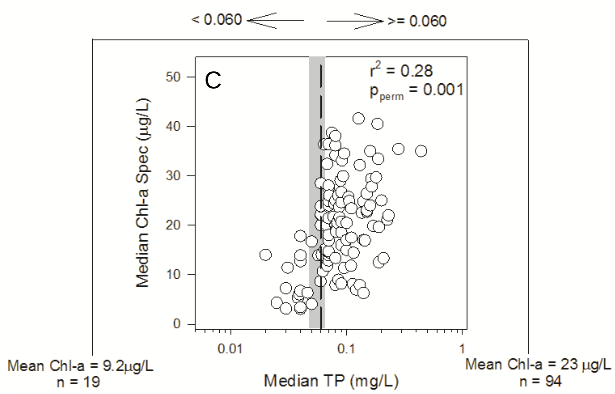
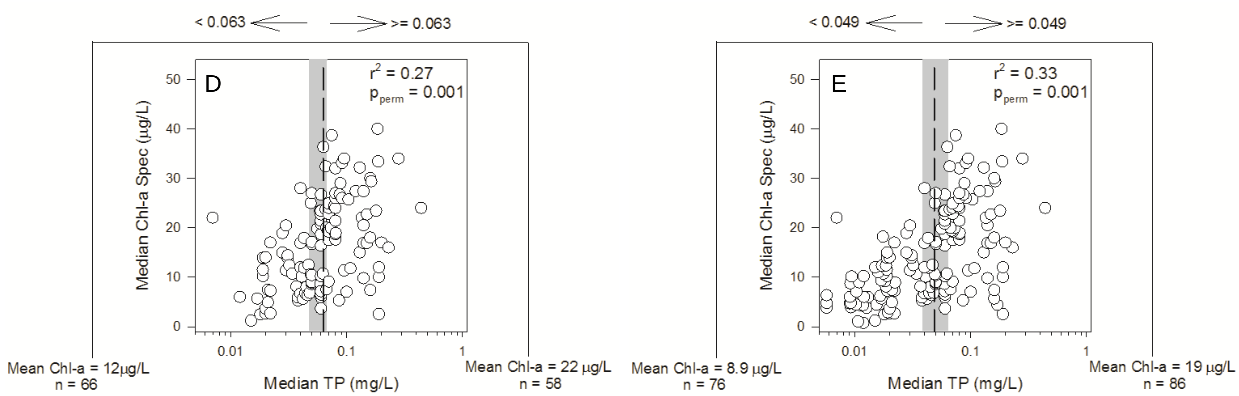
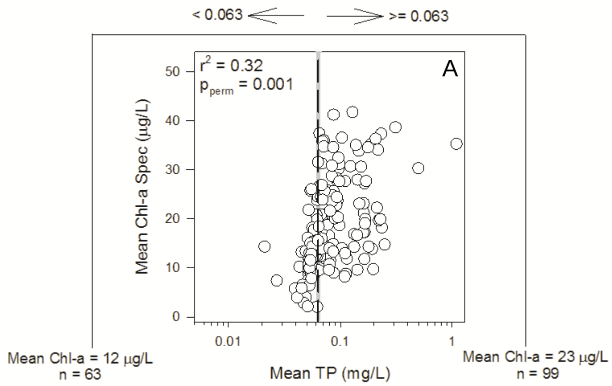
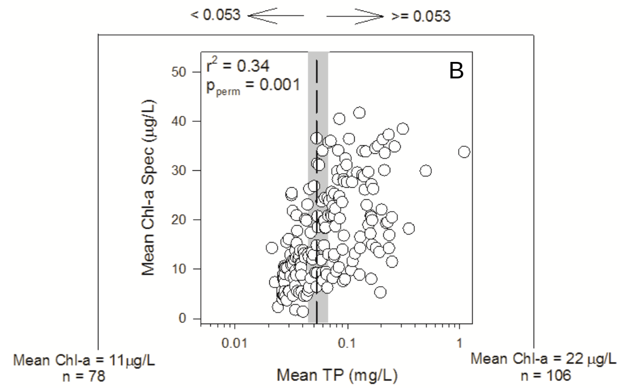


Figure 2.4.1. CART models of spectrophotometric chlorophyll-a vs. TP concentrations for Texas reservoirs. These models used medians from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-b), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of chl-a and/or TP could affect nutrient thresholds.







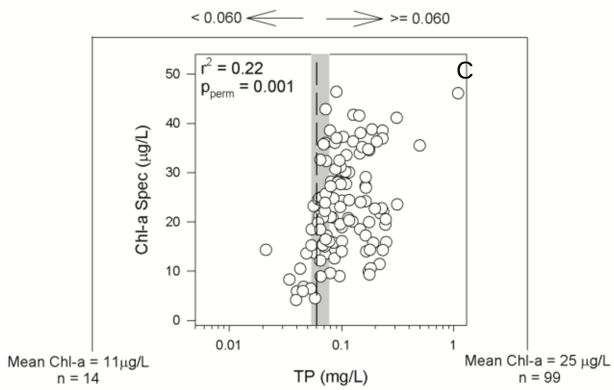
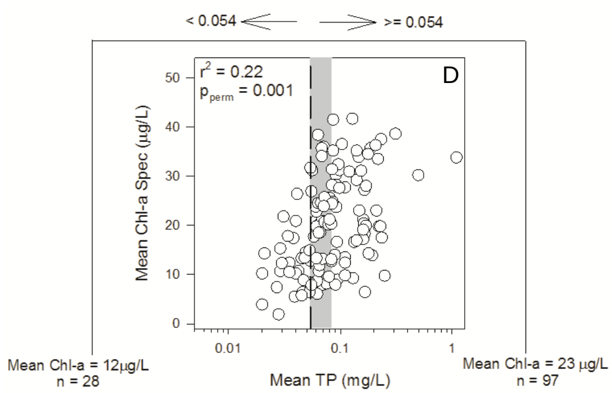
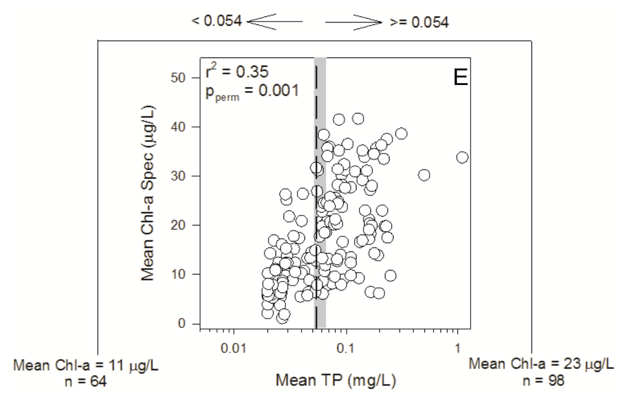
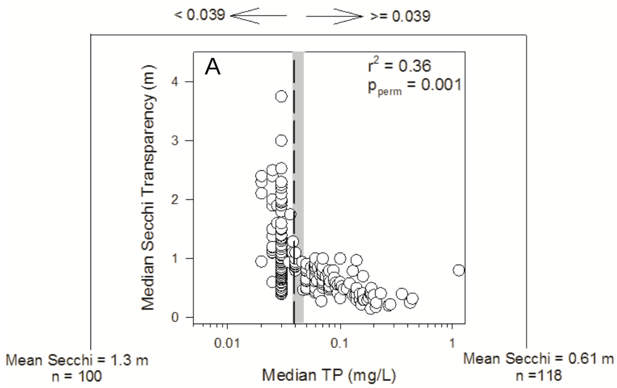
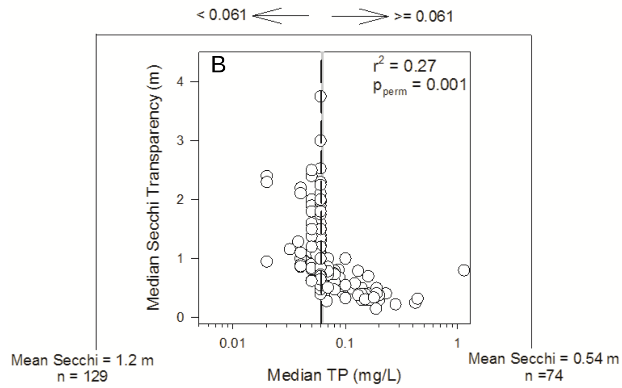


Figure 2.4.2. CART models of spectrophotometric chlorophyll-a vs. TP concentrations for Texas reservoirs. These models used means from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-b), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of chl-a and/or TP could affect nutrient thresholds.









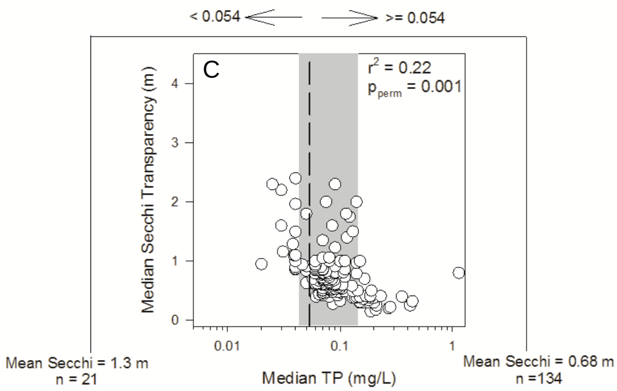
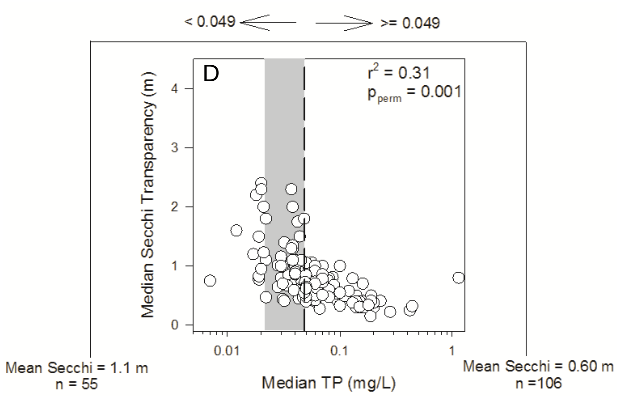
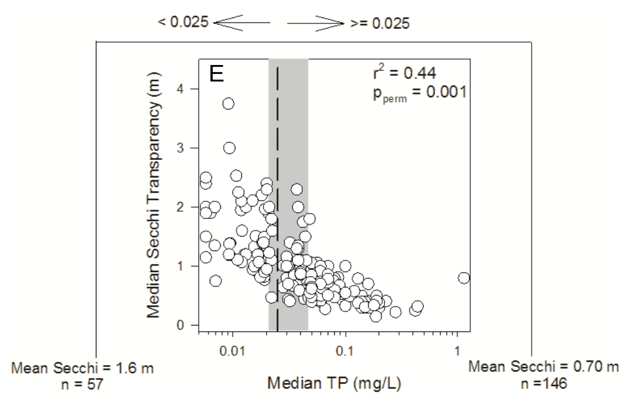
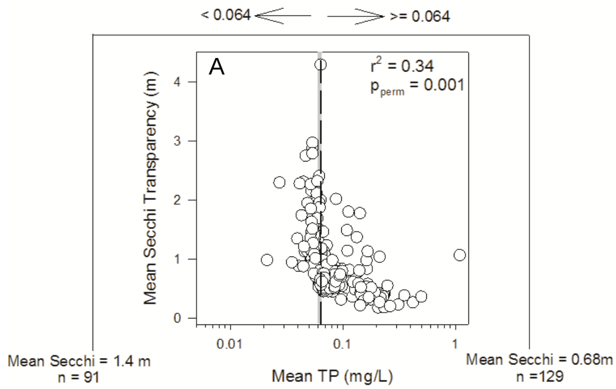
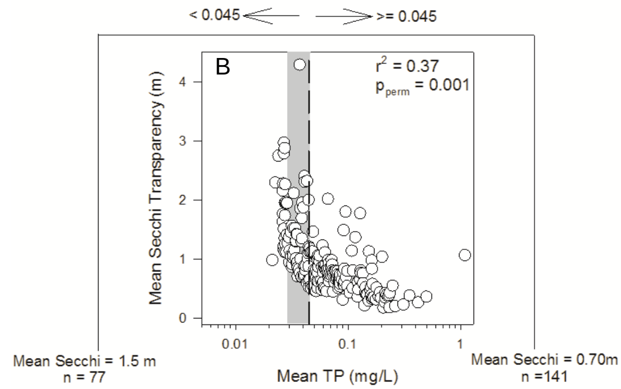


Figure 2.4.3. CART models of Secchi transparency vs. TP concentration for Texas reservoirs. These models used medians from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of TP could affect nutrient thresholds.









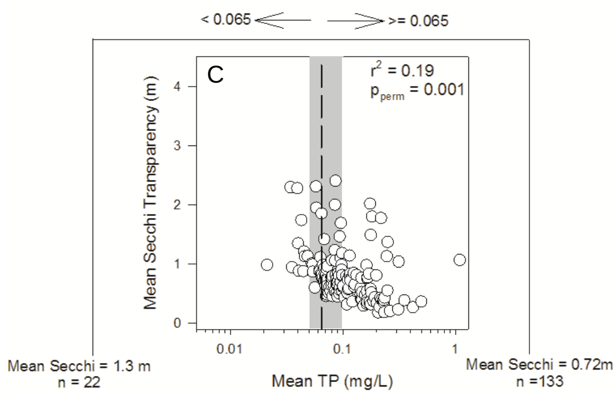
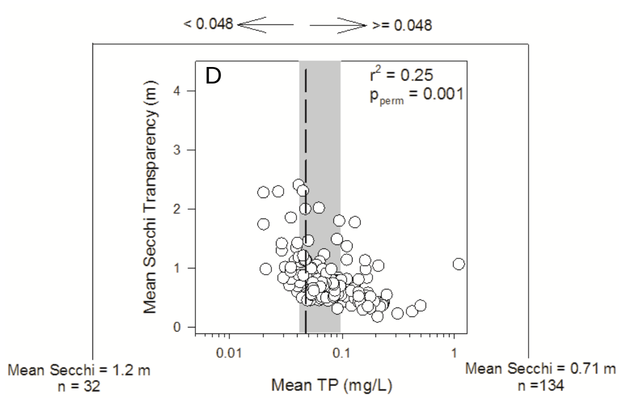


Figure 2.4.4. CART models of Secchi transparency vs. TP concentration for Texas reservoirs. These models used means from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of TP could affect nutrient thresholds.



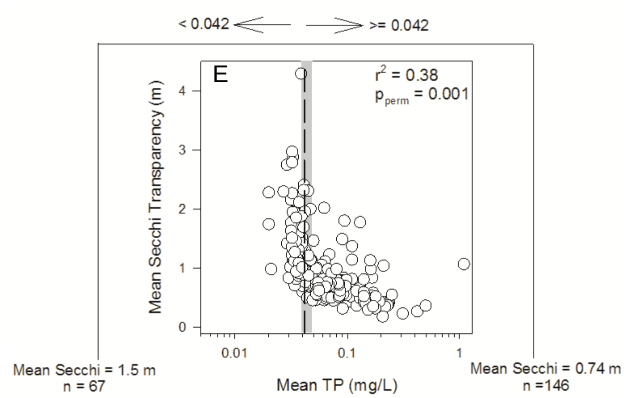


Table 2.4.2. Summary of TN thresholds identified in the 5 reservoir median/mean datasets.

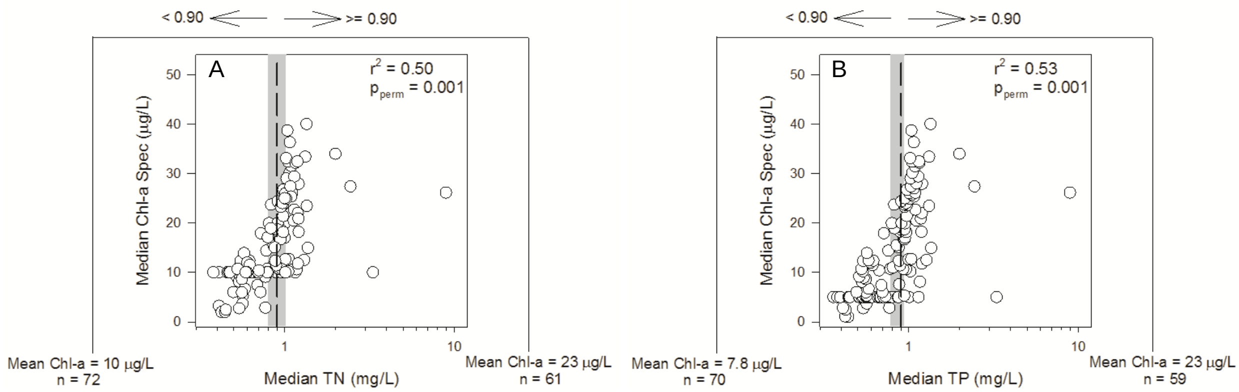
Dataset	Primary TN threshold (mg/L)			
	Median Chl-a	Median Secchi	Mean Chl-a	Mean Secchi
1) SubDL	0.90	0.60	0.88	0.81
2) Sub1/2DL	0.90	0.58	0.87	0.56
3) Deletion	0.79	0.58	0.80	0.64
4) Cen0-80%	0.90	NA	0.88	NA
5) Cen0-100%	0.90	NA	0.88	NA

Thresholds in median TN concentrations

A summary of TN thresholds for chl-a and Secchi transparency response is provided in Table 2.4.2. In contrast to the results of analyses of biological responses vs. TP, the TN thresholds identified in CART analyses did not vary greatly between the five datasets. Thresholds in TN relative to chl-a (Figs. 2.4.5-2.4.6), were slightly (approximately 25%) higher than thresholds in TN relative to Secchi transparency (Figs. 2.4.7-2.4.8), with the exception of the threshold identified for mean Secchi transparency vs. TN in Dataset 1, which was in range with chl-a thresholds.

As with TP thresholds, the TN thresholds identified in these analyses almost always exceeded the USEPA recommended criteria for TN concentrations in lakes and reservoirs located in aggregate ecoregions within Texas (0.36-0.78 mg/L TN), with the exception of aggregate ecoregion V-Cultivated Great Plains. The TN thresholds for Texas reservoirs were more similar to estimates of reference conditions based on NLA data, generally exceeding TN reference concentrations estimated for arid aggregate ecoregions II-Western Mountains and III-Xeric West, but lower than estimates for IV-Grass Plains lakes and for lakes and reservoirs in V-Cultivated Great Plains.

In general, CART models based on TN concentrations had good predictive power for both chl-a (r^2=0.45-0.54) and Secchi transparency (r^2=0.25-0.58), regardless of the measure of central tendency or approach to handling censored data. In fact, CART models of biological response variables vs. TN consistently had up to 2x greater predictive power than models of biological response variables vs. TP, suggesting that TN plays an important limiting role in biological productivity in Texas reservoirs. Therefore, it may be important to set numeric nutrient criteria for both TP and TN in Texas reservoirs.



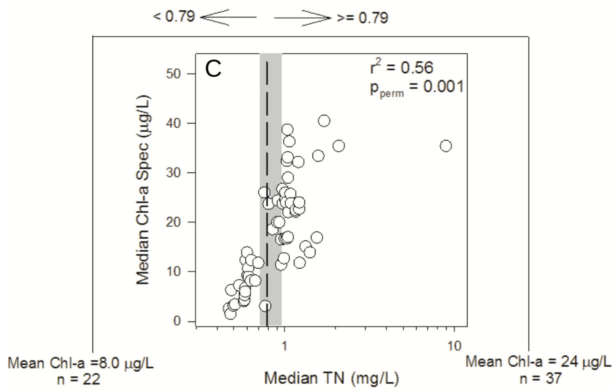
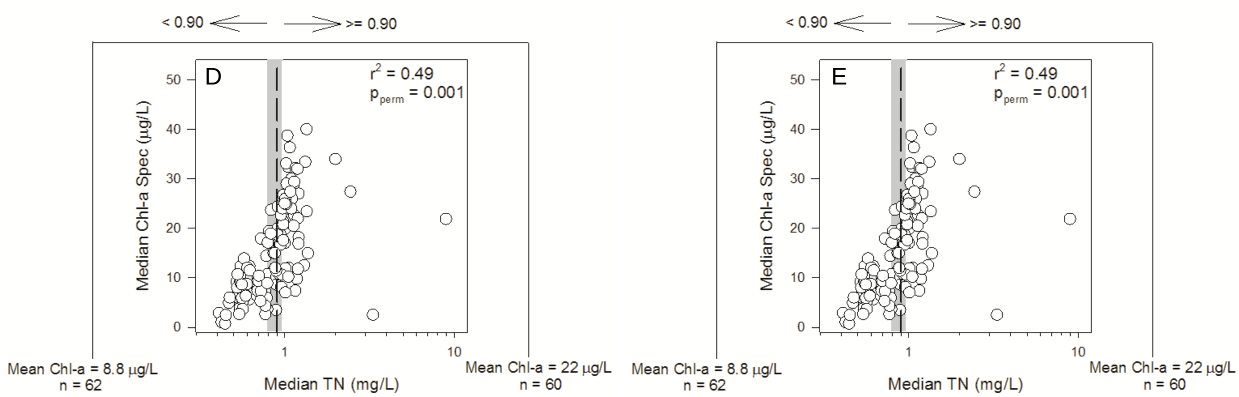
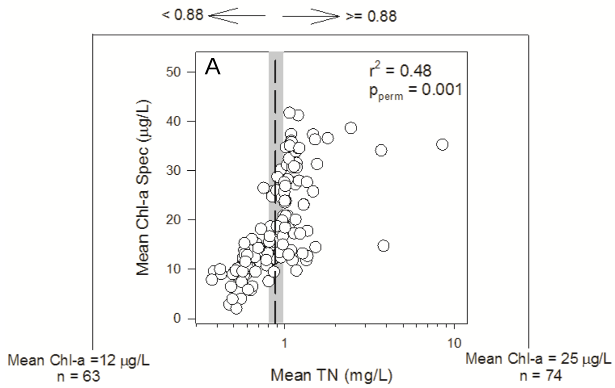
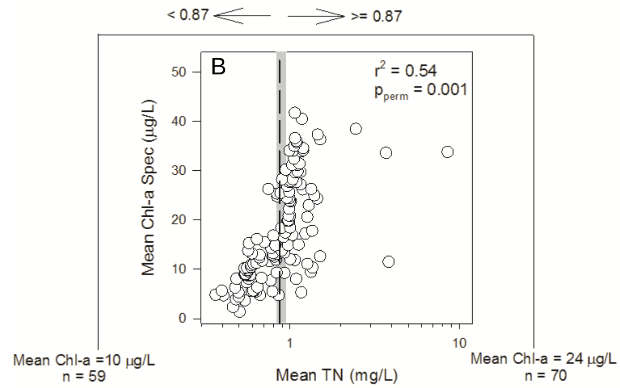


Figure 2.4.5. CART models of spectrophotometric chlorophyll-a vs. TN concentration for Texas reservoirs. These models used medians from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of chlorophyll-a could affect nutrient thresholds.







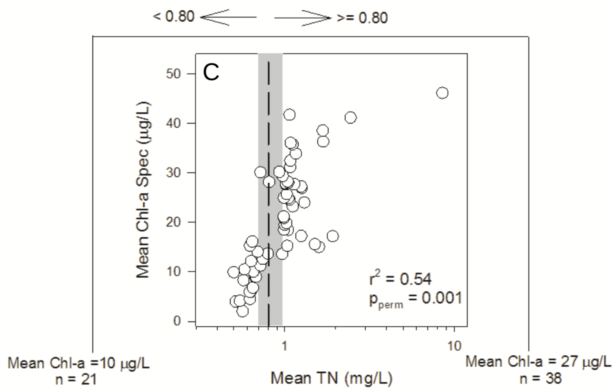
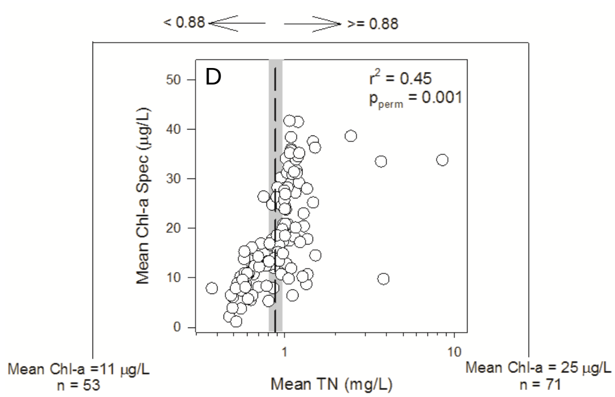
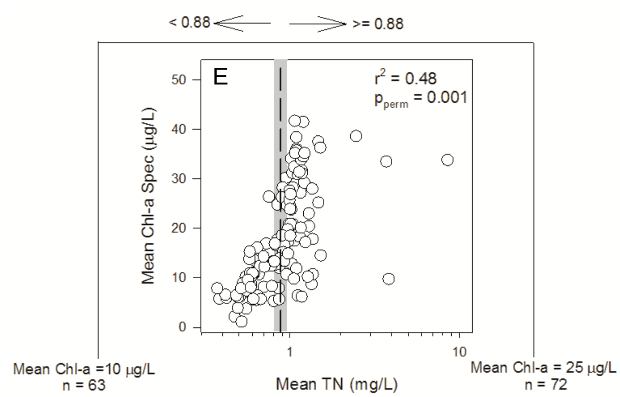
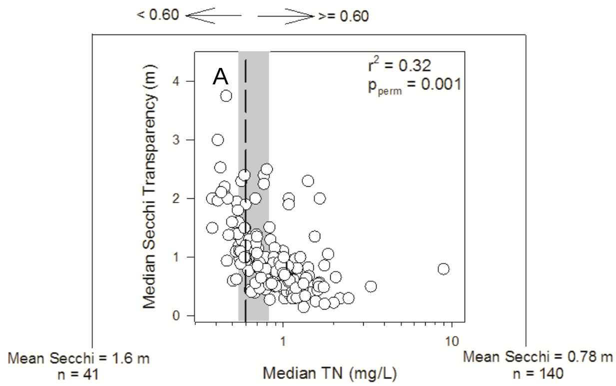
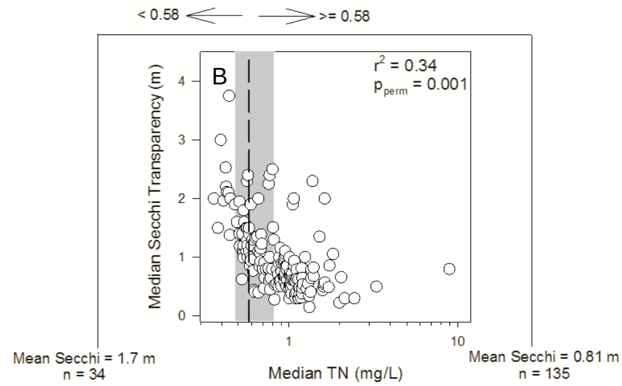


Figure 2.4.6. CART models of spectrophotometric chlorophyll-a vs. TN concentration for Texas reservoirs. These models used means from datasets 1-5 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), after deleting censored observations (C), or by using statistical techniques for estimating measures of central tendency in censored datasets (D-E). Peer-reviewed statistical techniques were only appropriate for datasets with 0-80% censoring (D). Therefore, a method was devised to estimate measures of central tendency for datsets with <80% censoring (E), allowing exploration of how data from stations with 80-100% censoring of chlorophyll-a could affect nutrient thresholds.









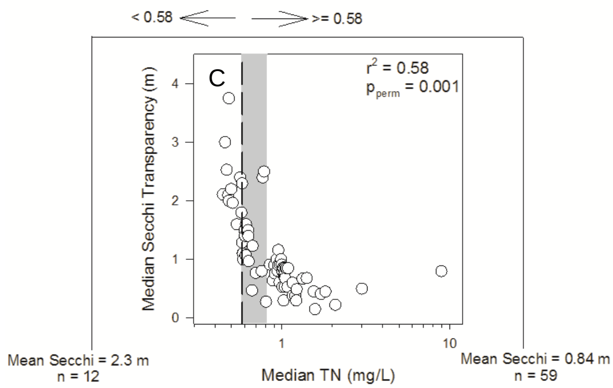
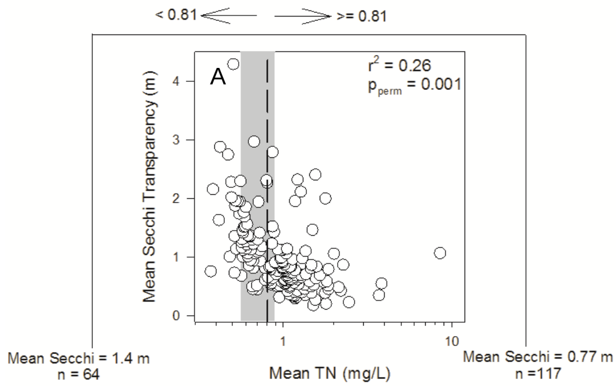
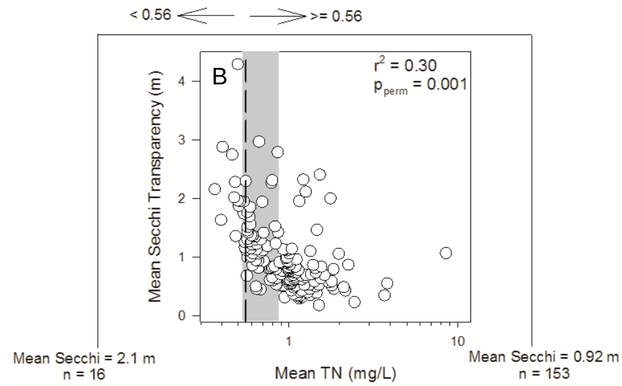


Figure 2.4.7. CART models of Secchi transparency vs. TN concentration for Texas reservoirs. These models used medians from datasets 1-3 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), or after deleting censored observations (C). Measures of central tendency using techniques for censored data analysis were not calculated for TN concentration or Secchi transparency because the frequency of censored observations was lower for these parameters than for TP or chlorophyll-a.





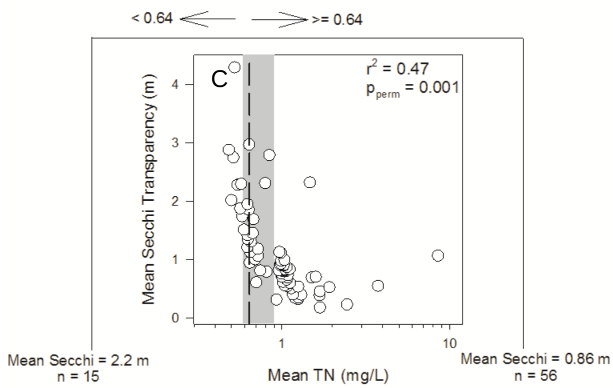


Figure 2.4.8. CART models of Secchi transparency vs. TN concentration for Texas reservoirs. These models used means from datasets 1-3 that were calculated after substituting the detection limit or half the detection limit for censored observations (A-B), or after deleting censored observations (C). Measures of central tendency using techniques for censored data analysis were not calculated for TN concentration or Secchi transparency because the frequency of censored observations was lower for these parameters than for TP or chlorophyll-a.

REFERENCES

Binkley, D., G.G. Ice, J. Kaye, and C.A. Williams. 2004. Nitrogen and phosphorus concentrations in forest streams of the United States. Journal of American Water Resources Association, 40:1277-1291.

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

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

East T L, Sharfstein B. 2006. Development of a decision tree model for the prediction of the limitation potential of phytoplankton in Lake Okeechobee, Florida, USA. Arch. Hydrobiol. 2006;165:127-144.

Evans-White, M.A., B.E. Haggard, and J.T. Scott. *In Press*. A review of stream nutrient criteria development in the United States. Jounal of Environmental Quality.

Flum, T., and S.C. Nodvin. 1995. Factors affecting streamwater chemistry in the Great Smoky Mountains, USA. Water, Air, and Soil Pollution, 85(3):1707-1712.

Longing, D, and B.E. Haggard. 2010. Distributions of median nutrient and chlorophyll concentrations across the Red River Basin, USA. Journal of Environmental Quality, 39(6): 1966-74.

Helsel, DR. 2012. Statistics for Censored Environmental Data Using MiniTab and R. 2nd Ed. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Herlihy, A.T., N.C. Kamman, J.C. Sifneos, D. Charles, M.D. Enache, and R.J. Stevenson. 2013. Using multiple approaches to develop nutrient criteria for lakes in the conterminous USA. Freshwater Science, 32(2):367-384.

Herlihy, A.T., and J.C. Sifneos. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous USA. Journal of the North American Benthological Society 27:932-948.

Ice, G. and D. Binkley. 2003. Forest stormwater concentrations of nitrogen and phosphorus: A comparison with EPA's proposed water quality criteria. Journal of Forestry. 101:21-28.

King RS, and CJ Richardson. 2003. Integrating bioasessment and ecological risk assessment: An approach to developing numerical water-quality critera. Environmental Management. 31:795-809. doi:10.1007/s00267-002-0036-4

King RS, Baker ME, Whigham DF, Weller DE, Jordan TE, Kazyak PF, Hurd MK. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecological Applications 15:137-153.

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(3):998-1008.

Qian SS, King RS, Richardson CJ. 2003. Two methods for the detection of environmental thresholds. Ecological Modeling 166:87-97.

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

Smith, R.A., R.B. Alexander, and G.E. Schwarz. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. Environmental Science and Technology, 37:3039-3047.

Suplee, M.W., A. Varghese, and J. Cleland. 2007. Developing nutrient criteria for streams: an evaluation of the frequency distribution method. Journal of the American Water Resources Association, 43:453-471.

Urban, D.L. (2002) "Classification and regression trees" In: McCune, B., Grace, J.B. eds. , Analysis of Ecological Communities, MjM Software Design, Gleneden Beach, Oregon, USA, pp 221-231.

USEPA (US Environmental Protection Agency). 2000. Nutreint criteria technical guidance manual: lakes and reservoirs. EPA-822-b00-001. Office of Water, US Environmental Protection Agency, Washington, DC.

Vaga, R.M., A.T. Herlihy, R. Miller, and M.M. Sytsma. 2006. A classification system of Pacific Northwest reservoirs with respect to nutrient processing. EPA 910-$-06-003. Region 10, US Environmental Protection Agency, Seattle, Washington.

APPENDICES

Appendix 2.1. Microsoft Excel macro used to add combined parameters to the Texas reservoirs water quality raw database.

```
Sub Add_Columns()
Dim i As Integer
For i = 3 To ThisWorkbook.Sheets.Count
Sheets(i).Activate

Columns(18).Insert Shift:=xlToRight
Columns(19).Insert Shift:=xlToRight
Columns(22).Insert Shift:=xlToRight
Columns(23).Insert Shift:=xlToRight
Columns(30).Insert Shift:=xlToRight
Columns(31).Insert Shift:=xlToRight
Columns(38).Insert Shift:=xlToRight
Columns(39).Insert Shift:=xlToRight
Columns(46).Insert Shift:=xlToRight
Columns(47).Insert Shift:=xlToRight
Columns(52).Insert Shift:=xlToRight
Columns(53).Insert Shift:=xlToRight
Columns(72).Insert Shift:=xlToRight
Columns(73).Insert Shift:=xlToRight
Columns(92).Insert Shift:=xlToRight
Columns(93).Insert Shift:=xlToRight
Columns(94).Insert Shift:=xlToRight
Columns(95).Insert Shift:=xlToRight
Columns(96).Insert Shift:=xlToRight
Columns(97).Insert Shift:=xlToRight
Columns(98).Insert Shift:=xlToRight
Columns(99).Insert Shift:=xlToRight
Columns(114).Insert Shift:=xlToRight
Columns(115).Insert Shift:=xlToRight
Columns(124).Insert Shift:=xlToRight
Columns(125).Insert Shift:=xlToRight
Columns(170).Insert Shift:=xlToRight
Columns(171).Insert Shift:=xlToRight
Columns(180).Insert Shift:=xlToRight
Columns(181).Insert Shift:=xlToRight
```

```
Columns(186).Insert Shift:=xlToRight
Columns(187).Insert Shift:=xlToRight

Cells(1, 18).Value = "00077m"
Cells(1, 19).Value = "Censor Code"
Cells(1, 22).Value = "00078C"
Cells(1, 23).Value = "Censor Code"
Cells(1, 30).Value = "00094C"
Cells(1, 31).Value = "Censor Code"
Cells(1, 38).Value = "00210C"
Cells(1, 39).Value = "Censor Code"
Cells(1, 46).Value = "00213C"
Cells(1, 47).Value = "Censor Code"
Cells(1, 52).Value = "00215C"
Cells(1, 53).Value = "Censor Code"
Cells(1, 72).Value = "00300C"
Cells(1, 73).Value = "Censor Code"
Cells(1, 92).Value = "00600A"
Cells(1, 93).Value = "Censor Code"
Cells(1, 94).Value = "00600B"
Cells(1, 95).Value = "Censor Code"
Cells(1, 96).Value = "00600i"
Cells(1, 97).Value = "Censor Code"
Cells(1, 98).Value = "00600C"
Cells(1, 99).Value = "Censor Code"
Cells(1, 114).Value = "00630C"
Cells(1, 115).Value = "Censor Code"
Cells(1, 124).Value = "00671C"
Cells(1, 125).Value = "Censor Code"
Cells(1, 170).Value = "82708C"
Cells(1, 171).Value = "Censor Code"
Cells(1, 180).Value = "89077m"
Cells(1, 181).Value = "Censor Code"
Cells(1, 186).Value = "89856C"
Cells(1, 187).Value = "Censor Code"

  Dim LastRow As Long
  LastRow = Cells(Rows.Count, "A").End(xlUp).Row
    If LastRow < 3 Then On Error Resume Next
```

```
'Runs equation for column 00077m
Range("r2").Select
ActiveCell.FormulaR1C1 = "=IF(rc[-2]="""","""",(rc[-2]/39.700787))"
Range("r2").Select
Selection.AutoFill Destination:=Range("R2:R" & LastRow)

'Runs equation for column 00077m Censor Code
Range("s2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]="""","""",rc[-2])"
Range("s2").Select
Selection.AutoFill Destination:=Range("s2:s" & LastRow)

'Runs equation for column 00078C
Range("v2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[-4]>0,rc[-4],if(rc[158]>0,rc[158],"""")))"
Range("v2").Select
Selection.AutoFill Destination:=Range("v2:v" & LastRow)

'Runs equation for column 00078C Censor Code
Range("w2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[-5]>0,rc[-4],if(rc[157]>0,rc[158],"""")))"
Range("w2").Select
Selection.AutoFill Destination:=Range("w2:w" & LastRow)

'Runs equation for column 00094C
Range("ad2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[2]>0,rc[2],""""))"
Range("ad2").Select
Selection.AutoFill Destination:=Range("ad2:ad" & LastRow)

'Runs equation for column 00094C Censor Code
Range("ae2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[1]>0,rc[2],""""))"
Range("ae2").Select
Selection.AutoFill Destination:=Range("ae2:ae" & LastRow)

'Runs equation for column 00210C
Range("al2").Select
```

```
ActiveCell.FormulaR1C1 = "=if(and(rc[-2]>0,rc[2]>0),(rc[-2]-rc[2]),"""")"
Range("al2").Select
Selection.AutoFill Destination:=Range("al2:al" & LastRow)

'Runs equation for column 00210C Censor Code
Range("am2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]="""","""",if(or(rc[-2]>0,rc[2]>0),3,0))"
Range("am2").Select
Selection.AutoFill Destination:=Range("am2:am" & LastRow)

'Runs equation for column 00213C
Range("at2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[-2]>0,rc[2]>0),(rc[-2]-rc[2]),"""")"
Range("at2").Select
Selection.AutoFill Destination:=Range("at2:at" & LastRow)

'Runs equation for column 00213C Censor Code
Range("au2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]="""","""",if(or(rc[-2]>0,rc[2]>0),3,0))"
Range("au2").Select
Selection.AutoFill Destination:=Range("au2:au" & LastRow)

'Runs equation for column 00215C
Range("az2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[-2]>0,rc[2]>0),(rc[-2]-rc[2]),"""")"
Range("az2").Select
Selection.AutoFill Destination:=Range("az2:az" & LastRow)

'Runs equation for column 00215C Censor Code
Range("ba2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]="""","""",if(or(rc[-2]>0,rc[2]>0),3,0))"
Range("ba2").Select
Selection.AutoFill Destination:=Range("ba2:ba" & LastRow)

'Runs equation for column 00400C
Range("bx2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[2]>0,rc[2],""""))"
Range("bx2").Select
Selection.AutoFill Destination:=Range("bx2:bx" & LastRow)
```

```
'Runs equation for column 00400C Censor Code
Range("by2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[1]>0,rc[2],""""))"
Range("by2").Select
Selection.AutoFill Destination:=Range("by2:by" & LastRow)

'Runs equation for column 00600A
Range("cn2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[18]>0,rc[20]>0),(rc[18]+rc[20]),0)"
Range("cn2").Select
Selection.AutoFill Destination:=Range("cn2:cn" & LastRow)

'Runs equation for column 00600A Censor Code
Range("co2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]>0,if(or(rc[18]>0,rc[20]>0),3,0),"""")"
Range("co2").Select
Selection.AutoFill Destination:=Range("co2:co" & LastRow)

'Runs equation for column 00600B
Range("cp2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[16]>0,rc[-4]>0),(rc[16]+rc[-4]),0)"
Range("cp2").Select
Selection.AutoFill Destination:=Range("cp2:cp" & LastRow)

'Runs equation for column 00600B Censor Code
Range("cq2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]>0,if(or(rc[16]>0,rc[-4]>0),3,0),"""")"
Range("cq2").Select
Selection.AutoFill Destination:=Range("cq2:cq" & LastRow)

'Runs equation for column 00600i
Range("cr2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[8]>0,rc[10]>0,rc[14]>0),(rc[8]+rc[10]+rc[14]),0)"
Range("cr2").Select
Selection.AutoFill Destination:=Range("cr2:cr" & LastRow)

'Runs equation for column 00600i Censor Code
Range("cs2").Select
```

```
ActiveCell.FormulaR1C1 = "=if(rc[-1]>0,if(or(rc[8]>0,rc[10]>0,rc[14]>0),3,0),"""")"
Range("cs2").Select
Selection.AutoFill Destination:=Range("cs2:cs" & LastRow)

'Runs equation for column 00600C
Range("ct2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-6]>0,rc[-6],if(rc[-4]>0,rc[-4],if(rc[-2]>0,rc[-2],"""")))"
Range("ct2").Select
Selection.AutoFill Destination:=Range("ct2:ct" & LastRow)

'Runs equation for column 00600C Censor Code
Range("cu2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-7]>0,rc[-6],if(rc[-5]>0,rc[-4],if(rc[-3]>0,rc[-2],"""")))"
Range("cu2").Select
Selection.AutoFill Destination:=Range("cu2:cu" & LastRow)

'Runs equation for column 00630C
Range("dj2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[-24]>0,rc[-24],""""))"
Range("dj2").Select
Selection.AutoFill Destination:=Range("dj2:dj" & LastRow)

'Runs equation for column 00630C Censor Code
Range("dk2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[-25]>0,rc[-24],""""))"
Range("dk2").Select
Selection.AutoFill Destination:=Range("dk2:dk" & LastRow)

'Runs equation for column 00671C
Range("dt2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[36]>0,rc[36],""""))"
Range("dt2").Select
Selection.AutoFill Destination:=Range("dt2:dt" & LastRow)

'Runs equation for column 00671C Censor Code
Range("du2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[35]>0,rc[36],""""))"
Range("du2").Select
Selection.AutoFill Destination:=Range("du2:du" & LastRow)
```

```
'Runs equation for column 82708C
Range("fn2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2],if(rc[2]>0,rc[2],""""))"
Range("fn2").Select
Selection.AutoFill Destination:=Range("fn2:fn" & LastRow)

'Runs equation for column 82708C Censor Code
Range("fo2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-3]>0,rc[-2],if(rc[1]>0,rc[2],""""))"
Range("fo2").Select
Selection.AutoFill Destination:=Range("fo2:fo" & LastRow)

'Runs equation for column 89077m
Range("fx2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]>0,rc[-2]*0.3048,"""")"
Range("fx2").Select
Selection.AutoFill Destination:=Range("fx2:fx" & LastRow)

'Runs equation for column 89077m Censor Code
Range("fy2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-2]="""","""",rc[-2])"
Selection.AutoFill Destination:=Range("fy2:fy" & LastRow)

'Runs equation for column 89856C
Range("gd2").Select
ActiveCell.FormulaR1C1 = "=if(and(rc[-2]>0,rc[-4]>0),(rc[-2]+rc[-4]),"""")"
Range("gd2").Select
Selection.AutoFill Destination:=Range("gd2:gd" & LastRow)

'Runs equation for column 89856C Censor Code
Range("ge2").Select
ActiveCell.FormulaR1C1 = "=if(rc[-1]="""","""",if(or(rc[-2]>0,rc[-4]>0),3,0))"
Range("ge2").Select
Selection.AutoFill Destination:=Range("ge2:ge" & LastRow)

'Replaces all zeros in 00600A, 00600B & 00600i with blank cell
  Dim rng As Range
  For Each rng In Range("cn2:cn" & LastRow)
```

```
        If rng.Value = 0 Then
            rng.Value = ""
        End If
    Next

    For Each rng In Range("cp2:cp" & LastRow)
        If rng.Value = 0 Then
            rng.Value = ""
        End If
    Next

    For Each rng In Range("cr2:cr" & LastRow)
        If rng.Value = 0 Then
            rng.Value = ""
        End If
    Next

Next

Next

End Sub
```

Appendix 2.2. Microsoft Excel macros used to calculate medians and counts for each parameters for each station in the Texas reservoirs water quality raw database.

```
Sub Add_Medians()

Dim i As Integer
For i = 4 To ThisWorkbook.Sheets.Count
Sheets(i).Activate

Dim LastRow As Long
Dim LastCol As Long
LastRow = Cells(Rows.Count, "a").End(xlUp).Row
LastCol = Cells(3, Columns.Count).End(xlToLeft).Column

Range("b4").Select
ActiveCell.Formula = "=Counta(r[3]C:R[" & LastRow & "]C)"
Range("B4").Select
Selection.AutoFill Destination:=Range(Cells(4, 2), Cells(4, LastCol))

Range("b5").Select
ActiveCell.Formula = "=Count(r[2]C:R[" & LastRow & "]C)"
Range("b5").Select
Selection.AutoFill Destination:=Range(Cells(5, 2), Cells(5, LastCol))

Range("b6").Select
ActiveCell.Formula = "=if(r[-1]C>0,median(r[1]C:R[" & LastRow & "]C),"""")"
Range("b6").Select
Selection.AutoFill Destination:=Range(Cells(6, 2), Cells(6, LastCol))

Next

End Sub
```

Appendix 2.3. Frequency distribution of median nutrient and chlorophyll-a concentrations from reservoirs among basin by level III ecoregions in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26- Southwestern Tablelands	4	0.050	--	0.058	0.060	0.060	--	0.060
2-25-High Plains	1	--	--	--	0.060	--	--	--
2-26-Southwestern Tablelands	3	0.055	--	--	0.060	--	--	1.140
2-27-Central Great Plains	10	0.050	0.050	0.143	0.160	0.196	0.211	0.220
2-29-Cross Timbers	3	0.060	--	--	0.073	--	--	0.080
2-33-East Central Texas Plains	3	0.050	--	--	0.060	--	--	0.067
3-35-South Central Plains	10	0.080	0.090	0.100	0.126	0.148	0.168	0.190
4-33-East Central Texas Plains	1	--	--	--	0.050	--	--	--
4-35-South Central Plains	13	0.040	0.050	0.050	0.055	0.060	0.082	0.100
5-33-East Central Texas Plains	8	0.060	0.060	0.060	0.060	0.073	0.085	0.095
5-35-South Central Plains	4	0.050	--	0.058	0.060	0.065	--	0.080
6-35-South Central Plains	31	0.050	0.050	0.060	0.060	0.060	0.100	0.190
8-29-Cross Timbers	42	0.020	0.040	0.051	0.061	0.080	0.095	0.155
8-32-Texas Blackland Prairies	18	0.015	0.019	0.021	0.050	0.060	0.090	0.188
8-33-East Central Texas Plains	13	0.032	0.042	0.060	0.078	0.164	0.218	0.720
8-35-South Central Plains	5	0.060	--	0.130	0.160	0.185	--	0.280
10-35-South Central Plains	17	0.100	0.156	0.170	0.200	0.280	0.366	0.440
12-25-High Plains	1	--	--	--	0.065	--	--	--
12-26-Southwestern Tablelands	1	--	--	--	0.060	--	--	--
12-27-Central Great Plains	1	--	--	--	0.060	--	--	--
12-29-Cross Timbers	44	0.020	0.022	0.050	0.060	0.060	0.070	0.080
12-30-Edwards Plateau	2	0.060	--	--	0.060	--	--	0.060
12-32-Texas Blackland Prairies	6	0.060	--	0.143	0.185	0.205	--	0.270
12-33-East Central Texas Plains	9	0.060	0.060	0.080	0.115	0.140	0.383	0.495
14-26-Southwestern Tablelands	3	0.060	--	--	0.060	--	--	0.085
14-27-Central Great Plains	12	0.050	0.060	0.060	0.060	0.060	0.060	0.140
14-29-Cross Timbers	3	0.060	--	--	0.060	--	--	0.060
14-30-Edwards Plateau	23	0.040	0.042	0.060	0.060	0.060	0.060	0.068
14-32-Texas Blackland Prairies	6	0.020	--	0.028	0.050	0.050	--	0.050
18-30-Edwards Plateau	2	0.050	--	--	0.055	--	--	0.060
21-31-Southern Texas Plains	5	0.057	--	0.070	0.140	0.166	--	0.208
23-24-Chihuahuan Deserts	5	0.050	--	0.050	0.050	0.060	--	0.060
23-31-Southern Texas Plains	1	--	--	--	0.060	--	--	--

Total Nitrogen (TN; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26- Southwestern Tablelands	4	0.47	--	0.50	0.57	0.65	--	0.69
2-25-High Plains	1	--	--	--	0.62	--	--	--
2-26-Southwestern Tablelands	3	0.53	--	--	0.54	--	--	8.94
2-27-Central Great Plains	4	0.51	--	0.66	0.80	1.21	--	2.18
2-29-Cross Timbers	1	--	--	--	0.67	--	--	--

2-33-East Central Texas Plains	2	0.74	--	--	0.77	--	--	0.79
3-35-South Central Plains	8	1.04	1.05	1.08	1.09	1.11	1.15	1.21
4-33-East Central Texas Plains	1	--	--	--	0.59	--	--	--
4-35-South Central Plains	12	0.53	0.55	0.60	0.70	0.77	0.79	1.01
5-33-East Central Texas Plains	7	0.83	0.92	0.99	1.00	1.12	1.14	1.14
5-35-South Central Plains	4	0.53	--	0.88	1.04	1.10	--	1.19
6-35-South Central Plains	18	0.38	0.44	0.56	0.74	1.01	1.33	3.32
8-29-Cross Timbers	37	0.47	0.56	0.69	0.91	1.01	1.06	1.20
8-32-Texas Blackland Prairies	12	0.36	0.48	0.76	0.93	1.00	1.18	1.32
8-33-East Central Texas Plains	13	0.88	0.89	0.95	1.03	1.16	1.65	6.59
8-35-South Central Plains	5	0.60	--	0.86	1.10	1.35	--	1.99
10-35-South Central Plains	1	--	--	--	1.16	--	--	--
12-25-High Plains	1	--	--	--	2.05	--	--	--
12-26-Southwestern Tablelands	1	--	--	--	0.90	--	--	--
12-27-Central Great Plains	1	--	--	--	0.82	--	--	--
12-29-Cross Timbers	37	0.53	0.62	0.70	0.86	1.18	1.64	1.85
12-30-Edwards Plateau	2	1.08	--	--	1.17	--	--	1.27
12-32-Texas Blackland Prairies	6	1.31	--	1.40	1.58	1.68	--	1.76
12-33-East Central Texas Plains	9	1.18	1.20	1.36	1.42	1.62	1.78	1.93
14-26-Southwestern Tablelands	1	--	--	--	1.98	--	--	--
14-27-Central Great Plains	12	0.67	0.70	0.76	0.96	1.11	1.36	2.45
14-29-Cross Timbers	3	0.60	--	--	0.64	--	--	0.65
14-30-Edwards Plateau	18	0.31	0.42	0.45	0.55	0.62	0.73	0.96
14-32-Texas Blackland Prairies	4	0.50	--	0.53	0.55	0.62	--	0.77
18-30-Edwards Plateau	1	--	--	--	0.41	--	--	--
21-31-Southern Texas Plains	3	0.49	--	--	0.99	--	--	1.08
23-24-Chihuahuan Deserts	5	0.46	--	0.59	0.80	1.21	--	1.63
23-31-Southern Texas Plains	1	--	--	--	3.11	--	--	--

Nitrite Plus Nitrate-Nitrogen (NO_X-N; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26- Southwestern Tablelands	4	0.04	--	0.04	0.04	0.04	--	0.05
2-25-High Plains	1	--	--	--	0.05	--	--	--
2-26-Southwestern Tablelands	3	0.04	--	--	0.04	--	--	5.70
2-27-Central Great Plains	4	0.04	--	0.04	0.04	0.04	--	0.05
2-29-Cross Timbers	3	0.04	--	--	0.04	--	--	0.06
2-33-East Central Texas Plains	3	0.05	--	--	0.05	--	--	0.12
3-35-South Central Plains	8	0.04	0.04	0.04	0.05	0.05	0.05	0.05
4-33-East Central Texas Plains	1	--	--	--	0.05	--	--	--
4-35-South Central Plains	12	0.02	0.04	0.04	0.04	0.05	0.05	0.05
5-33-East Central Texas Plains	6	0.04	--	0.04	0.04	0.06	--	0.15
5-35-South Central Plains	3	0.04	--	--	0.04	--	--	0.05
6-35-South Central Plains	31	0.04	0.04	0.05	0.05	0.07	0.11	1.94
8-29-Cross Timbers	37	0.00	0.01	0.01	0.03	0.04	0.07	0.32
8-32-Texas Blackland Prairies	12	0.00	0.01	0.07	0.08	0.10	0.11	0.31
8-33-East Central Texas Plains	13	0.01	0.01	0.02	0.04	0.18	0.20	4.47
8-35-South Central Plains	5	0.04	--	0.05	0.14	0.28	--	0.65
10-35-South Central Plains	1	--	--	--	0.19	--	--	--

12-25-High Plains	1	--	--	--	0.51	--	--	--
12-26-Southwestern Tablelands	1	--	--	--	0.04	--	--	--
12-27-Central Great Plains	1	--	--	--	0.04	--	--	--
12-29-Cross Timbers	32	0.01	0.03	0.04	0.05	0.07	0.13	0.33
12-30-Edwards Plateau	0	--	--	--	--	--	--	--
12-32-Texas Blackland Prairies	5	0.04	--	0.05	0.05	0.05	--	0.05
12-33-East Central Texas Plains	5	0.04	--	0.04	0.05	0.05	--	0.14
14-26-Southwestern Tablelands	3	0.02	--	--	0.02	--	--	0.02
14-27-Central Great Plains	12	0.02	0.04	0.04	0.05	0.05	0.07	0.16
14-29-Cross Timbers	3	0.04	--	--	0.05	--	--	0.06
14-30-Edwards Plateau	18	0.02	0.02	0.03	0.04	0.10	0.14	0.18
14-32-Texas Blackland Prairies	4	0.20	--	0.20	0.22	0.26	--	0.32
18-30-Edwards Plateau	2	0.07	--	--	0.08	--	--	0.10
21-31-Southern Texas Plains	5	0.02	--	0.02	0.02	0.02	--	0.03
23-24-Chihuahuan Deserts	5	0.04	--	0.14	0.15	0.24	--	0.44
23-31-Southern Texas Plains	1	--	--	--	0.05	--	--	--

Ortho-Phosphate (PO_4-P; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
1-26- Southwestern Tablelands	4	0.040	--	0.040	0.040	0.045	--	0.060
2-25-High Plains	1	--	--	--	0.060	--	--	--
2-26-Southwestern Tablelands	3	0.040	--	--	0.040	--	--	0.910
2-27-Central Great Plains	10	0.040	0.058	0.069	0.100	0.118	0.122	0.135
2-29-Cross Timbers	3	0.040	--	--	0.040	--	--	0.040
2-33-East Central Texas Plains	3	0.040	--	--	0.060	--	--	0.060
3-35-South Central Plains	10	0.007	0.008	0.025	0.060	0.060	0.060	0.060
4-33-East Central Texas Plains	1	--	--	--	0.040	--	--	--
4-35-South Central Plains	16	0.010	0.020	0.020	0.055	0.060	0.060	0.060
5-33-East Central Texas Plains	8	0.010	0.031	0.040	0.040	0.040	0.043	0.050
5-35-South Central Plains	4	0.040	--	0.040	0.040	0.040	--	0.040
6-35-South Central Plains	26	0.040	0.040	0.040	0.040	0.040	0.060	0.060
8-29-Cross Timbers	44	0.002	0.005	0.007	0.010	0.020	0.030	0.040
8-32-Texas Blackland Prairies	25	0.006	0.008	0.010	0.020	0.020	0.040	0.060
8-33-East Central Texas Plains	14	0.008	0.009	0.010	0.010	0.053	0.060	0.455
8-35-South Central Plains	10	0.040	0.040	0.063	0.080	0.089	0.117	0.130
10-35-South Central Plains	14	0.040	0.042	0.053	0.080	0.115	0.204	0.345
12-25-High Plains	1	--	--	--	0.040	--	--	--
12-26-Southwestern Tablelands	1	--	--	--	0.050	--	--	--
12-27-Central Great Plains	3	0.010	--	--	0.040	--	--	0.050
12-29-Cross Timbers	64	0.002	0.006	0.012	0.040	0.040	0.040	0.060
12-30-Edwards Plateau	2	0.040	--	--	0.040	--	--	0.040
12-32-Texas Blackland Prairies	6	0.040	--	0.060	0.062	0.098	--	0.120
12-33-East Central Texas Plains	9	0.040	0.040	0.040	0.040	0.060	0.218	0.270
14-26-Southwestern Tablelands	3	0.020	--	--	0.020	--	--	0.060
14-27-Central Great Plains	12	0.040	0.040	0.040	0.060	0.060	0.060	0.060
14-29-Cross Timbers	3	0.040	--	--	0.040	--	--	0.040
14-30-Edwards Plateau	23	0.010	0.010	0.019	0.030	0.040	0.040	0.040

14-32-Texas Blackland Prairies	6	0.010	--	0.013	0.020	0.020	--	0.020
18-30-Edwards Plateau	1	--	--	--	0.040	--	--	--
21-31-Southern Texas Plains	4	0.010	--	0.048	0.066	0.075	--	0.083
23-24-Chihuahuan Deserts	5	0.040	--	0.040	0.040	0.040	--	0.050
23-31-Southern Texas Plains	1	--	--	--	0.040	--	--	--

Chlorophyll-a (Chl-a; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
1-26- Southwestern Tablelands	3	5.30	--	--	5.89	--	--	19.0
2-25-High Plains	1	--	--	--	7.75	--	--	--
2-26-Southwestern Tablelands	3	3.68	--	--	12.5	--	--	65.3
2-27-Central Great Plains	3	9.20	--	--	14.2	--	--	82.9
2-29-Cross Timbers	3	9.55	--	--	21.1	--	--	23.4
2-33-East Central Texas Plains	2	7.35	--	--	16.6	--	--	25.9
3-35-South Central Plains	8	16.2	18.9	26.1	29.5	34.2	45.6	71.1
4-33-East Central Texas Plains	1	--	--	--	5.97	--	--	--
4-35-South Central Plains	8	5.60	6.91	8.73	17.5	20.2	23.0	28.9
5-33-East Central Texas Plains	1	--	--	--	25.9	--	--	--
5-35-South Central Plains	2	44.5	--	--	47.1	--	--	49.6
6-35-South Central Plains	19	3.11	4.30	10.2	13.0	33.9	39.4	52.4
8-29-Cross Timbers	4	4.57	--	10.0	14.3	20.0	--	29.7
8-32-Texas Blackland Prairies	4	10.5	--	14.5	16.8	20.9	--	30.7
8-33-East Central Texas Plains	0	--	--	--	--	--	--	--
8-35-South Central Plains	1	--	--	--	8.95	--	--	--
10-35-South Central Plains	4	6.78	--	9.08	13.1	20.7	--	33.6
12-25-High Plains	1	--	--	--	45.5	--	--	--
12-26-Southwestern Tablelands	1	--	--	--	13.6	--	--	--
12-27-Central Great Plains	0	--	--	--	--	--	--	--
12-29-Cross Timbers	38	3.00	3.60	10.4	16.5	22.0	25.4	29.5
12-30-Edwards Plateau	2	3.50	--	--	4.55	--	--	5.60
12-32-Texas Blackland Prairies	2	9.10	--	--	23.7	--	--	38.2
12-33-East Central Texas Plains	6	14.7	--	20.7	22.2	35.1	--	69.4
14-26-Southwestern Tablelands	2	16.2	--	--	26.2	--	--	36.1
14-27-Central Great Plains	11	3.44	5.00	6.01	10.0	15.2	18.0	53.3
14-29-Cross Timbers	3	7.59	--	--	8.80	--	--	9.70
14-30-Edwards Plateau	22	0.130	0.286	3.31	5.92	8.68	12.5	15.7
14-32-Texas Blackland Prairies	4	0.250	--	1.26	1.95	3.64	--	7.68
18-30-Edwards Plateau	1	--	--	--	3.00	--	--	--
21-31-Southern Texas Plains	4	5.35	--	12.0	15.2	16.5	--	17.1
23-24-Chihuahuan Deserts	5	3.00	--	3.00	3.05	21.9	--	31.8
23-31-Southern Texas Plains	0	--	--	--	--	--	--	--

Appendix 2.4. Frequency distribution of median nutrient and chlorophyll-a concentrations from reservoirs among level IV Ecoregions in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	7	0.020	0.020	0.023	0.025	0.055	0.064	0.070
25i-Llano Estacado	2	0.060	--	--	0.060	--	--	0.060
26a-Canadian/Cimarron Breaks	2	0.050	--	--	0.055	--	--	0.060
26b-Flat Tablelands & Valleys	1	--	--	--	0.060	--	--	--
26c-Caprock Canyon/Bdlnd/Brk	8	0.020	0.041	0.058	0.060	0.063	0.111	0.208
27h-Red Prairie	5	0.060	--	0.060	0.075	0.080	--	0.100
27i-Broken Red Plains	8	0.050	0.057	0.060	0.060	0.065	0.086	0.100
27j-Limestone Plains	2	0.040	--	--	0.051	--	--	0.063
29b-Eastern Cross Timbers	17	0.040	0.046	0.060	0.060	0.060	0.073	0.100
29c-Western Cross Timbers	22	0.035	0.050	0.060	0.068	0.093	0.184	0.220
29d-Grand Prairie	25	0.020	0.044	0.060	0.060	0.095	0.140	0.230
29e-Limestone Cut Plain	14	0.050	0.063	0.099	0.162	0.189	0.215	0.280
29f-Carbonate Cross Timbers	8	0.015	0.019	0.020	0.040	0.050	0.056	0.070
30a-Edwards Plateau Woodland	0	--	--	--	--	--	--	--
30b-Llano Uplift	12	0.040	0.047	0.059	0.060	0.080	0.081	0.160
30c-Balcones Canyonlands	10	0.020	0.038	0.050	0.060	0.060	0.190	0.280
30d-Semiarid Edwards Plateau	0	--	--	--	--	--	--	--
31c-Texas-Tamaulipan Thrnsrcb	7	0.020	0.023	0.043	0.060	0.165	0.182	0.200
31d-Rio Grande Fldpln/Terrace	0	--	--	--	--	--	--	--
32a-Northern Blackland Prairie	33	0.040	0.050	0.060	0.070	0.150	0.206	0.720
33a-Northern Post Oak Savanna	26	0.038	0.050	0.050	0.060	0.088	0.173	0.420
33b-Southern Post Oak Savanna	16	0.050	0.050	0.060	0.060	0.117	0.260	0.440
33f-Floodplains & Low Terrace	2	0.060	--	--	0.060	--	--	0.060
35a-Tertiary Uplands	29	0.040	0.054	0.060	0.060	0.070	0.168	1.140
35b-Floodplains & Low Terrace	11	0.040	0.055	0.060	0.060	0.078	0.150	0.355
35e-Southern Tertiary Uplands	26	0.015	0.026	0.060	0.060	0.066	0.118	0.160
35f-Flatwoods	17	0.025	0.046	0.060	0.060	0.070	0.104	0.200

Total Nitrogen (TN; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	5	0.54	--	0.59	0.73	0.74	--	0.82
25i-Llano Estacado	2	0.62	--	--	0.63	--	--	0.63
26a-Canadian/Cimarron Breaks	2	1.21	--	--	1.42	--	--	1.63
26b-Flat Tablelands & Valleys	1	--	--	--	0.46	--	--	--
26c-Caprock Canyon/Bdlnd/Brk	7	0.41	0.42	0.45	0.59	0.69	0.85	0.99
27h-Red Prairie	4	0.52	--	0.55	0.71	0.89	--	1.01
27i-Broken Red Plains	8	0.51	0.56	0.59	0.63	0.82	1.24	1.34
27j-Limestone Plains	2	0.53	--	--	0.54	--	--	0.56
29b-Eastern Cross Timbers	14	0.41	0.46	0.53	0.61	0.81	1.00	1.10
29c-Western Cross Timbers	15	0.64	0.67	0.69	0.81	0.98	1.06	1.20
29d-Grand Prairie	18	0.36	0.59	0.94	1.07	1.29	1.48	3.32

29e-Limestone Cut Plain	12	0.55	0.97	1.02	1.15	1.24	1.72	1.99
29f-Carbonate Cross Timbers	4	0.38	--	0.52	0.68	0.80	--	0.81
30a-Edwards Plateau Woodland	0	--	--	--	--	--	--	--
30b-Llano Uplift	11	0.47	0.58	0.87	0.98	1.08	1.16	1.18
30c-Balcones Canyonlands	4	0.31	--	0.52	0.62	0.69	--	0.79
30d-Semiarid Edwards Plateau	0	--	--	--	--	--	--	--
31c-Texas-Tamaulipan Thrnsrcb	3	0.53	--	--	0.66	--	--	0.71
31d-Rio Grande Fldpln/Terrace	0	--	--	--	--	--	--	--
32a-Northern Blackland Prairie	26	0.45	0.52	0.59	1.09	1.56	1.70	6.59
33a-Northern Post Oak Savanna	24	0.38	0.56	0.68	0.89	1.08	1.59	2.18
33b-Southern Post Oak Savanna	11	0.54	0.60	0.62	0.96	1.02	1.32	1.71
33f-Floodplains & Low Terrace	2	1.06	--	--	1.46	--	--	1.85
35a-Tertiary Uplands	18	0.60	0.69	0.80	0.91	1.11	1.80	8.94
35b-Floodplains & Low Terrace	10	0.53	0.54	0.72	0.90	1.33	1.52	1.98
35e-Southern Tertiary Uplands	21	0.60	0.73	0.79	0.95	0.99	1.14	2.45
35f-Flatwoods	14	0.43	0.53	0.71	0.94	1.12	1.34	3.11

Ortho-Phosphate (PO_4-P; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	11	0.006	0.006	0.010	0.040	0.040	0.040	0.040
25i-Llano Estacado	2	0.040	--	--	0.040	--	--	0.040
26a-Canadian/Cimarron Breaks	2	0.040	--	--	0.045	--	--	0.050
26b-Flat Tablelands & Valleys	1	--	--	--	0.040	--	--	--
26c-Caprock Canyon/Bdlnd/Brk	8	0.013	0.018	0.020	0.038	0.045	0.064	0.072
27h-Red Prairie	8	0.002	0.004	0.007	0.030	0.043	0.050	0.050
27i-Broken Red Plains	8	0.040	0.040	0.040	0.040	0.043	0.053	0.060
27j-Limestone Plains	3	0.007	--	--	0.008	--	--	0.040
29b-Eastern Cross Timbers	17	0.005	0.016	0.030	0.040	0.040	0.060	0.060
29c-Western Cross Timbers	27	0.006	0.009	0.010	0.020	0.060	0.120	0.135
29d-Grand Prairie	28	0.006	0.009	0.010	0.040	0.040	0.060	0.100
29e-Limestone Cut Plain	15	0.010	0.010	0.012	0.040	0.050	0.072	0.115
29f-Carbonate Cross Timbers	9	0.002	0.002	0.006	0.010	0.019	0.024	0.040
30a-Edwards Plateau Woodland	0	--	--	--	--	--	--	--
30b-Llano Uplift	12	0.005	0.005	0.007	0.009	0.040	0.040	0.080
30c-Balcones Canyonlands	12	0.010	0.010	0.014	0.020	0.040	0.040	0.100
30d-Semiarid Edwards Plateau	0	--	--	--	--	--	--	--
31c-Texas-Tamaulipan Thrnsrcb	9	0.006	0.006	0.040	0.060	0.080	0.081	0.087
31d-Rio Grande Fldpln/Terrace	0	--	--	--	--	--	--	--
32a-Northern Blackland Prairie	36	0.003	0.010	0.020	0.040	0.040	0.067	0.455
33a-Northern Post Oak Savanna	27	0.005	0.032	0.040	0.040	0.060	0.060	0.240
33b-Southern Post Oak Savanna	16	0.010	0.020	0.040	0.040	0.045	0.090	0.345
33f-Floodplains & Low Terrace	2	0.040	--	--	0.050	--	--	0.060
35a-Tertiary Uplands	29	0.020	0.020	0.040	0.040	0.060	0.100	0.910
35b-Floodplains & Low Terrace	15	0.013	0.020	0.020	0.040	0.045	0.060	0.205
35e-Southern Tertiary Uplands	27	0.006	0.008	0.010	0.040	0.060	0.060	0.110
35f-Flatwoods	16	0.003	0.003	0.018	0.040	0.040	0.050	0.083

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	5	0.01	--	0.04	0.05	0.05	--	0.05
25i-Llano Estacado	2	0.04	--	--	0.04	--	--	0.04
26a-Canadian/Cimarron Breaks	2	0.04	--	--	0.09	--	--	0.15
26b-Flat Tablelands & Valleys	1	--	--	--	0.14	--	--	--
26c-Caprock Canyon/Bdlnd/Brk	8	0.02	0.02	0.03	0.05	0.11	0.17	0.24
27h-Red Prairie	4	0.01	--	0.02	0.03	0.05	--	0.06
27i-Broken Red Plains	8	0.02	0.03	0.04	0.04	0.04	0.10	0.20
27j-Limestone Plains	2	0.01	--	--	0.02	--	--	0.03
29b-Eastern Cross Timbers	14	0.02	0.02	0.02	0.05	0.06	0.12	0.15
29c-Western Cross Timbers	17	0.01	0.01	0.03	0.04	0.05	0.05	0.26
29d-Grand Prairie	17	0.04	0.04	0.05	0.05	0.07	0.18	1.94
29e-Limestone Cut Plain	13	0.01	0.01	0.02	0.04	0.19	0.31	0.65
29f-Carbonate Cross Timbers	4	0.00	--	0.00	0.03	0.09	--	0.20
30a-Edwards Plateau Woodland	0	--	--	--	--	--	--	--
30b-Llano Uplift	11	0.01	0.01	0.03	0.04	0.08	0.19	0.23
30c-Balcones Canyonlands	4	0.02	--	0.03	0.04	0.04	--	0.04
30d-Semiarid Edwards Plateau	0	--	--	--	--	--	--	--
31c-Texas-Tamaulipan Thrnsrcb	3	0.01	--	--	0.04	--	--	0.04
31d-Rio Grande Fldpln/Terrace	0	--	--	--	--	--	--	--
32a-Northern Blackland Prairie	22	0.01	0.01	0.04	0.05	0.08	0.20	4.47
33a-Northern Post Oak Savanna	21	0.04	0.04	0.04	0.05	0.05	0.14	0.51
33b-Southern Post Oak Savanna	12	0.02	0.04	0.04	0.06	0.15	0.24	0.33
33f-Floodplains & Low Terrace	1	--	--	--	0.05	--	--	--
35a-Tertiary Uplands	25	0.02	0.04	0.05	0.05	0.06	0.10	5.70
35b-Floodplains & Low Terrace	11	0.00	0.02	0.02	0.04	0.05	0.07	0.14
35e-Southern Tertiary Uplands	22	0.03	0.04	0.04	0.05	0.11	0.18	0.32
35f-Flatwoods	16	0.00	0.02	0.05	0.06	0.09	0.12	0.23

Chlorophyll-a (Chl-a; µg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24a-Chihuahuan Basins & Playas	4	5.97	--	12.93	15.8	17.8	--	22.1
25i-Llano Estacado	2	3.76	--	--	5.33	--	--	6.90
26a-Canadian/Cimarron Breaks	2	21.9	--	--	26.8	--	--	31.8
26b-Flat Tablelands & Valleys	1	--	--	--	3.00	--	--	--
26c-Caprock Canyon/Bdlnd/Brk	5	3.00	--	5.00	5.00	5.89	--	17.1
27h-Red Prairie	5	20.4	--	21.7	23.3	23.9	--	29.1
27i-Broken Red Plains	7	5.30	9.62	12.8	13.2	26.9	40.8	49.6
27j-Limestone Plains	0	--	--	--	--	--	--	--
29b-Eastern Cross Timbers	11	2.75	3.00	6.75	8.80	11.4	15.7	29.7
29c-Western Cross Timbers	8	9.55	10.6	14.7	18.5	21.6	23.2	23.4
29d-Grand Prairie	13	0.130	1.62	8.80	21.2	34.0	39.0	43.4
29e-Limestone Cut Plain	3	4.35	--	--	12.3	--	--	28.7
29f-Carbonate Cross Timbers	2	1.60	--	--	2.86	--	--	4.12

30a-Edwards Plateau Woodland	0	--	--	--	--	--	--	--
30b-Llano Uplift	4	3.00	--	13.3	18.4	23.4	--	33.6
30c-Balcones Canyonlands	4	0.220	--	0.220	0.758	3.21	--	8.95
30d-Semiarid Edwards Plateau	0	--	--	--	--	--	--	--
31c-Texas-Tamaulipan Thrnsrcb	1	--	--	--	3.00	--	--	--
31d-Rio Grande Fldpln/Terrace	0	--	--	--	--	--	--	--
32a-Northern Blackland Prairie	19	5.00	5.57	7.43	9.10	13.1	28.2	69.4
33a-Northern Post Oak Savanna	17	3.00	3.09	5.60	9.26	29.5	55.7	82.9
33b-Southern Post Oak Savanna	11	0.250	2.30	5.40	12.2	16.9	33.0	52.4
33f-Floodplains & Low Terrace	1	--	--	--	18.2	--	--	--
35a-Tertiary Uplands	15	9.15	10.2	13.9	16.2	24.8	33.6	65.3
35b-Floodplains & Low Terrace	7	3.68	4.61	8.87	16.2	18.3	22.4	28.1
35e-Southern Tertiary Uplands	15	3.44	5.71	7.90	14.1	25.9	41.6	53.3
35f-Flatwoods	12	4.26	5.04	7.53	15.2	18.2	24.7	29.5

Appendix 2.4. CART and nCPA analyses of total nutrients (TP and TN) and biological response variables (chl-a and Secchi transparency) vs. geospatial parameters

ANALYSIS: TP VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TP ~ DEV + AG + DEVAG + FOR + WET, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=225 (539 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.1498404      0 1.0000000 1.0096601 0.5858348
2 0.0100000      2 0.7003193 0.9181988 0.5922306

Node number 1: 225 observations,    complexity param=0.1498404
  mean=0.08818667, MSE=0.008533928
  left son=2 (189 obs) right son=3 (36 obs)
  Primary splits:
      DEV   < 0.129617  to the left,  improve=0.12525000, (0 missing)
      AG    < 0.7956235 to the left,  improve=0.07839621, (0 missing)
      WET   < 0.1022505 to the left,  improve=0.04755659, (0 missing)
      FOR   < 0.411452  to the right, improve=0.03851963, (0 missing)
      DEVAG < 0.4185645 to the left,  improve=0.03349267, (0 missing)

Node number 2: 189 observations
  mean=0.07391799, MSE=0.001447983

Node number 3: 36 observations,    complexity param=0.1498404
  mean=0.1630972, MSE=0.03905468
  left son=6 (20 obs) right son=7 (16 obs)
  Primary splits:
      DEV   < 0.1814565 to the right, improve=0.238220400, (0 missing)
      AG    < 0.1569475 to the left,  improve=0.155922700, (0 missing)
      WET   < 0.128063  to the left,  improve=0.055252180, (0 missing)
      DEVAG < 0.6454605 to the right, improve=0.002982233, (0 missing)
      FOR   < 0.2531865 to the left,  improve=0.002982233, (0 missing)

Node number 6: 20 observations
  mean=0.076825, MSE=0.004781357

Node number 7: 16 observations
  mean=0.2709375, MSE=0.06096318
```

ANALYSIS: TP VS. % DEVELOPED (nCPA)

```
        cp      r2  mean left mean right pperm       5%      25%      50%       75%       95%
[1,] 0.129617 0.12525 0.07391799  0.1630972 0.005 0.1017385 0.129617 0.129617 0.1513275 0.1634705
```

ANALYSIS: TN VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TN ~ DEV + AG + DEVAG + FOR + WET, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=181 (583 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.09961010      0 1.0000000 1.010165 0.6723097
2 0.04008549      1 0.9003899 1.068159 0.6730557

Node number 1: 181 observations,    complexity param=0.0996101
  mean=0.993553, MSE=0.5241426
  left son=2 (170 obs) right son=3 (11 obs)
  Primary splits:
      AG    < 0.7956235 to the left,  improve=0.09961010, (0 missing)
      FOR   < 0.088972  to the right, improve=0.08193365, (0 missing)
      DEVAG < 0.478796  to the left,  improve=0.05724520, (0 missing)
      WET   < 0.0010565 to the left,  improve=0.02825559, (0 missing)
      DEV   < 0.0822125 to the left,  improve=0.02245882, (0 missing)

Node number 2: 170 observations
  mean=0.93543, MSE=0.1678483

Node number 3: 11 observations
  mean=1.891818, MSE=5.17141
```

ANALYSIS: TN VS. %AGRICULTURE (nCPA)

```
     cp         r2 mean left mean right pperm       5%      25%
[1,] 0.5734735 0.06590911 0.8757029   1.152727 0.015 0.2864075 0.292038
          50%       75%      95%
[1,] 0.3812195 0.5734735 0.764093
```

ANALYSIS: CHL-A SPEC VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = CHLASPEC ~ DEV + AG + DEVAG + FOR + WET, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=188 (576 observations deleted due to missingness)
```

```
        CP nsplit rel error    xerror      xstd
1 0.20306306      0 1.0000000 1.0114633 0.10137397
2 0.11021306      1 0.7969369 0.8693148 0.09348357
3 0.05818845      2 0.6867239 0.7668073 0.09428029

Node number 1: 188 observations,    complexity param=0.2030631
  mean=15.74423, MSE=70.41278
  left son=2 (84 obs) right son=3 (104 obs)
  Primary splits:
      AG    < 0.3812195 to the left,  improve=0.20306310, (0 missing)
      FOR   < 0.270093  to the right, improve=0.15277210, (0 missing)
      DEVAG < 0.439641  to the left,  improve=0.13640140, (0 missing)
      DEV   < 0.0836725 to the left,  improve=0.08987070, (0 missing)
      WET   < 0.0013345 to the left,  improve=0.04001766, (0 missing)

Node number 2: 84 observations
  mean=11.53679, MSE=36.23632

Node number 3: 104 observations,    complexity param=0.1102131
  mean=19.14255, MSE=72.17003
  left son=6 (58 obs) right son=7 (46 obs)
  Primary splits:
      DEV   < 0.0805365 to the left,  improve=0.19438030, (0 missing)
      FOR   < 0.172714  to the right, improve=0.09020111, (0 missing)
      WET   < 0.025696  to the left,  improve=0.07903534, (0 missing)
      AG    < 0.5155905 to the right, improve=0.07696095, (0 missing)
      DEVAG < 0.714559  to the left,  improve=0.05165058, (0 missing)

Node number 6: 58 observations
  mean=15.80698, MSE=55.12649

Node number 7: 46 observations
  mean=23.34826, MSE=61.94327
```

ANALYSIS: CHL-A SPEC VS. %AGRICULTURE (nCPA)

```
          cp        r2 mean left mean right pperm       5%       25%      50%       75%       95%

[1,] 0.3812195 0.2014543  11.55530   19.16238 0.001 0.1807155 0.2229245 0.380723 0.3812195
0.3984258
```

ANALYSIS: CHL-A FLUORO VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ DEV + AG + DEVAG + FOR + WET, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=107 (657 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.20062530      0 1.0000000 1.0199234 0.1945914
2 0.08366002      1 0.7993747 0.8948692 0.1958856

Node number 1: 107 observations,    complexity param=0.2006253
  mean=16.6815, MSE=157.9041
  left son=2 (33 obs) right son=3 (74 obs)
  Primary splits:
      AG    < 0.216106  to the left,  improve=0.20062530, (0 missing)
      FOR   < 0.659483  to the right, improve=0.15452850, (0 missing)
      DEV   < 0.0282865 to the left,  improve=0.13568510, (0 missing)
      DEVAG < 0.4185645 to the left,  improve=0.12462920, (0 missing)
      WET   < 0.0026105 to the left,  improve=0.08712593, (0 missing)

Node number 2: 33 observations
  mean=8.25303, MSE=18.66373

Node number 3: 74 observations
  mean=20.44014, MSE=174.1909
```

ANALYSIS: CHL-A FLUORO VS. %AGRICULTURE (nCPA)

```
           cp        r2 mean left mean right pperm       5%      25%      50%       75%      95%

[1,] 0.216106 0.1978861   8.25303   20.23569 0.002 0.214601 0.216106 0.216106 0.3763425 0.380723
```

ANALYSIS: SECCHI TRANSPARENCY VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = SECCHI ~ DEV + AG + DEVAG + FOR + WET, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=280 (484 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.19676044      0 1.0000000 1.0083370 0.1470082
2 0.05985253      1 0.8032396 0.9074662 0.1105638
3 0.05094173      3 0.6835345 0.8958089 0.1280595

Node number 1: 280 observations,    complexity param=0.1967604
  mean=0.8750659, MSE=0.2842033
```

```
  left son=2 (237 obs) right son=3 (43 obs)
  Primary splits:
      AG    < 0.1807155 to the right, improve=0.19676040, (0 missing)
      DEVAG < 0.4158975 to the right, improve=0.13617100, (0 missing)
      FOR   < 0.460874  to the left,  improve=0.12767790, (0 missing)
      DEV   < 0.2106645 to the left,  improve=0.08894405, (0 missing)
      WET   < 0.0024925 to the right, improve=0.06495087, (0 missing)

Node number 2: 237 observations,    complexity param=0.05985253
  mean=0.7743394, MSE=0.164762
  left son=4 (49 obs) right son=5 (188 obs)
  Primary splits:
      WET   < 0.1022505 to the right, improve=0.12039000, (0 missing)
      DEV   < 0.0329345 to the right, improve=0.11901770, (0 missing)
      DEVAG < 0.4158975 to the right, improve=0.09682778, (0 missing)
      FOR   < 0.461346  to the left,  improve=0.09390616, (0 missing)
      AG    < 0.237692  to the right, improve=0.03882942, (0 missing)

Node number 3: 43 observations
  mean=1.430233, MSE=0.5783895

Node number 4: 49 observations
  mean=0.4984694, MSE=0.02658796

Node number 5: 188 observations,    complexity param=0.05985253
  mean=0.8462417, MSE=0.1757698
  left son=10 (102 obs) right son=11 (86 obs)
  Primary splits:
      DEVAG < 0.6495895 to the right, improve=0.14600530, (0 missing)
      FOR   < 0.3476055 to the left,  improve=0.13818250, (0 missing)
      AG    < 0.6333615 to the right, improve=0.11603280, (0 missing)
      DEV   < 0.0329345 to the right, improve=0.09387643, (0 missing)
      WET   < 0.0024925 to the right, improve=0.06794835, (0 missing)

Node number 10: 102 observations
  mean=0.6991441, MSE=0.07992529

Node number 11: 86 observations
  mean=1.020706, MSE=0.2333446
```

ANALYSIS: SECCHI TRANSPARENCY VS. %AGRICULTURE (nCPA)

```
        cp      r2 mean left mean right pperm     5%     25%     50%    75%    95%
[1,] 0.1807155 0.1987103  1.430233  0.7712809 0.001 0.046897 0.0713125 0.1807155 0.22472
0.249951
```

ANALYSIS: TP VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (CART)

```
Call:
mvpart(form = TP ~ FLOW, data = res, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=225 (539 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.14325984      0 1.0000000 1.0213159 0.5910329
2 0.01027322      1 0.8567402 0.8738878 0.5936996
3 0.01000000      4 0.8259205 0.9115226 0.5999339

Node number 1: 225 observations,    complexity param=0.1432598
  mean=0.08818667, MSE=0.008533928
  left son=2 (211 obs) right son=3 (14 obs)
  Primary splits:
      FLOW < 146.5354 to the left,  improve=0.1432598, (0 missing)

Node number 2: 211 observations,    complexity param=0.01027322
  mean=0.07918009, MSE=0.007110649
  left son=4 (152 obs) right son=5 (59 obs)
  Primary splits:
      FLOW < 0.5705   to the right, improve=0.01306824, (0 missing)

Node number 3: 14 observations
  mean=0.2239286, MSE=0.01033635

Node number 4: 152 observations,    complexity param=0.01027322
  mean=0.07317434, MSE=0.001148756
  left son=8 (89 obs) right son=9 (63 obs)
  Primary splits:
      FLOW < 9.046    to the left,  improve=0.07738761, (0 missing)

Node number 5: 59 observations
  mean=0.09465254, MSE=0.02213778

Node number 8: 89 observations
  mean=0.06524157, MSE=0.0002950287

Node number 9: 63 observations,    complexity param=0.01027322
```

```
 mean=0.08438095, MSE=0.002140327
 left son=18 (42 obs) right son=19 (21 obs)
 Primary splits:
    FLOW < 12.77955 to the right, improve=0.1932525, (0 missing)

Node number 18: 42 observations
 mean=0.07, MSE=0.00125566

Node number 19: 21 observations
 mean=0.1131429, MSE=0.002668789
```

ANALYSIS: TP VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
        cp       r2 mean left mean right pperm      5%     25%      50%      75%      95%
[1,] 146.5354 0.1432598 0.0791801  0.2239286 0.001 55.6975 102.303 146.5354 146.5354 146.5354
```

ANALYSIS: TP VS. AREA-WEIGHTED PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (CART)

```
mvpart(form = TP ~ WFLOW, data = res, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=224 (540 observations deleted due to missingness)

        CP nsplit rel error    xerror      xstd
1 0.1087299      0 1.0000000 1.0082574 0.5886408
2 0.0100000      1 0.8912701 0.9180467 0.5955364

Node number 1: 224 observations,    complexity param=0.1087299
 mean=0.08795536, MSE=0.008559988
 left son=2 (207 obs) right son=3 (17 obs)
 Primary splits:
    WFLOW < 0.02802327 to the left,  improve=0.1087299, (0 missing)

Node number 2: 207 observations
 mean=0.07921256, MSE=0.007218649

Node number 3: 17 observations
 mean=0.1944118, MSE=0.01262907
```

ANALYSIS: TP VS. AREA-WEIGHTED PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
        r2  mean left mean right pperm         5%        25%
[1,] 0.01749716 0.06538549 0.07930583  0.1869444 0.003 8.72455e-06 0.01581333
          50%        75%        95%
[1,] 0.01749716 0.01749716 0.02802327
```

ANALYSIS: TN VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (CART)

```
Call:
mvpart(form = TN ~ FLOW, data = res, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=181 (583 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.01657323      0 1.0000000 1.016471 0.6748876
2 0.01517214      1 0.9834268 1.031295 0.6849714

Node number 1: 181 observations,    complexity param=0.01657323
  mean=0.993553, MSE=0.5241426
  left son=2 (142 obs) right son=3 (39 obs)
  Primary splits:
     FLOW < 11.2913 to the left,  improve=0.01657323, (0 missing)

Node number 2: 142 observations
  mean=0.9447084, MSE=0.5830165

Node number 3: 39 observations
  mean=1.171397, MSE=0.2694661
```

ANALYSIS: TN VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
          cp         r2 mean left mean right pperm     5%      25%  50%     75%
[1,] 11.2913 0.01657323 0.9447084   1.171397 0.425 0.1605 0.934275 9.54 11.2913
          95%
[1,] 28.40785
```

ANALYSIS: TN VS. AREA-WEIGHTED PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (CART)

```
Call:
mvpart(form = TN ~ WFLOW, data = res, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=180 (584 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.02560534      0 1.0000000 1.007852 0.6711150
2 0.01643063      1 0.9743947 1.035624 0.7057096

Node number 1: 180 observations,    complexity param=0.02560534
  mean=0.9925172, MSE=0.5268603
  left son=2 (77 obs) right son=3 (103 obs)
  Primary splits:
```

```
    WFLOW < 0.0005641535 to the left,  improve=0.02560534, (0 missing)

Node number 2: 77 observations
  mean=0.8581831, MSE=0.9440398

Node number 3: 103 observations
  mean=1.092942, MSE=0.1914128
```

ANALYSIS: CHL-A SPEC VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE BOTH UNWEIGHTED AND WEIGHTED BY WATERSHED AREA (CART)

```
Call:
mvpart(form = CHLASPEC ~ FLOW + WFLOW, data = res, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
  n=188 (576 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.16179966      0 1.0000000 1.0102987 0.1017639
2 0.04909363      1 0.8382003 0.8857347 0.1011393

Node number 1: 188 observations,    complexity param=0.1617997
  mean=15.74423, MSE=70.41278
  left son=2 (83 obs) right son=3 (104 obs), 1 observation remains
  Primary splits:
      WFLOW < 0.0005795715 to the left,  improve=0.1597899, (1 missing)
      FLOW  < 6.051        to the left,  improve=0.1293861, (0 missing)

Node number 2: 83 observations
  mean=12.00699, MSE=31.6449

Node number 3: 104 observations
  mean=18.7763, MSE=81.43496
```

ANALYSIS: CHL-A SPEC VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
        cp        r2 mean left mean right pperm     5%   25%     50%   75%      95%
[1,] 6.051 0.1280203   12.9723   19.01843 0.001 2.3005 6.051 6.10001 9.186 11.31735
```

ANALYSIS: CHL-A SPEC VS. AREA-WEIGHTED PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
               cp        r2 mean left mean right pperm          5%          25%         50%         75%         95%
[1,] 0.0005795715 0.1640093  12.00699   18.87877 0.001 0.000575457 0.0005795715 0.001379370
0.001406015 0.004360364
```

ANALYSIS: CHL-A FLUORO VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE BOTH UNWEIGHTED AND WEIGHTED BY WATERSHED AREA (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ FLOW + WFLOW, data = res, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
 n=107 (657 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.08919433      0 1.0000000 1.0232478 0.1968084
2 0.03174253      2 0.8216113 0.9928186 0.1958336

Node number 1: 107 observations,    complexity param=0.08919433
  mean=16.6815, MSE=157.9041
  left son=2 (78 obs) right son=3 (29 obs)
  Primary splits:
      FLOW  < 10.0325     to the left,  improve=0.04084958, (0 missing)
      WFLOW < 0.000649621 to the left,  improve=0.02757663, (0 missing)

Node number 2: 78 observations
  mean=15.13288, MSE=129.1935

Node number 3: 29 observations,    complexity param=0.08919433
  mean=20.84672, MSE=211.3263
  left son=6 (13 obs) right son=7 (16 obs)
  Primary splits:
      FLOW  < 16.1942     to the right, improve=0.3791857, (0 missing)
      WFLOW < 0.002832678 to the left,  improve=0.0808809, (0 missing)

Node number 6: 13 observations
  mean=10.91577, MSE=66.46771

Node number 7: 16 observations
  mean=28.91562, MSE=183.7848
```

ANALYSIS: CHL-A FLUORO VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
          cp         r2 mean left mean right pperm     5%    25%     50%     75%      95%
[1,] 10.0325 0.03270359  15.10604   20.21982 0.388 0.6105 9.6325 10.0325 16.1942 17.94605
```

ANALYSIS: SECCHI TRANSPARENCY VS. PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE BOTH UNWEIGHTED AND WEIGHTED BY WATERSHED AREA (CART)

```
Call:
mvpart(form = SECCHI ~ FLOW + WFLOW, data = res, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=280 (484 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.09978950      0 1.0000000 1.0053436 0.1462437
2 0.05693196      1 0.9002105 0.9900078 0.1394106

Node number 1: 280 observations,    complexity param=0.0997895
  mean=0.8750659, MSE=0.2842033
  left son=2 (159 obs) right son=3 (120 obs), 1 observation remains
  Primary splits:
      WFLOW < 0.001333546 to the right, improve=0.09967584, (1 missing)
      FLOW  < 146.5354    to the right, improve=0.08307091, (0 missing)

Node number 2: 159 observations
  mean=0.7282455, MSE=0.1488586

Node number 3: 120 observations
  mean=1.068812, MSE=0.3997289
```

ANALYSIS: SECCHI TRANSPARENCY VS. AREA-WEIGHTED PERMITTED WASTEWATER TREATMENT PLANT DISCHARGE (nCPA)

```
          cp        r2 mean left mean right pperm          5%         25%         50%        75%        95%

[1,] 0.001333546 0.1013934   1.06939  0.7242374 0.001 4.537995e-05 0.001333546 0.001357162
0.01665524 0.02125464
```

ANALYSIS: TP VS. REGIONS (BASIN, ECOREGION III, BASIN-ECOREGION III)

```
Call:
mvpart(form = TP ~ BASIN + ECO3 + BASECO3, data = res, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
 n=226 (538 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.27110097      0  1.000000 1.005476 0.5829068
2 0.03434798      1  0.728899 1.211352 0.6032979

Node number 1: 226 observations,    complexity param=0.271101
  mean=0.0885708, MSE=0.008529368
  left son=2 (204 obs) right son=3 (22 obs)
  Primary splits:
```

```
    BASECO3 splits as  LRLL-LLRLLLLLLLLRLLLLLLLLLLLLLLLR, improve=0.27110100, (0 missing)
    BASIN   splits as  LRLLLLRLLLLLLL, improve=0.16280710, (0 missing)
    ECO3    splits as  LLRLLLRRLR, improve=0.06038331, (0 missing)

Node number 2: 204 observations
  mean=0.07277941, MSE=0.001521827

Node number 3: 22 observations
  mean=0.235, MSE=0.04975455
```

ANALYSIS: TN VS. REGIONS (BASIN, ECOREGION III, BASIN-ECOREGION III)

```
Call:
mvpart(form = TN ~ BASIN + ECO3 + BASECO3, data = res, xval = 10,
    method = "anova", minsplit = 20, minbucket = 10)
  n=182 (582 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.13904626      0 1.0000000 1.007687 0.6717830
2 0.04613713      1 0.8609537 1.188952 0.6736048

Node number 1: 182 observations,    complexity param=0.1390463
  mean=0.9936983, MSE=0.5212666
  left son=2 (167 obs) right son=3 (15 obs)
  Primary splits:
    BASECO3 splits as  LLRL-LLRR-LLLLLLRLLL-L-LLLLLLLLL, improve=0.13904630, (0 missing)
    BASIN   splits as  LRRL-LR-LRLLLL, improve=0.07594053, (0 missing)
    ECO3    splits as  LRRRLL-LLL, improve=0.04939417, (0 missing)

Node number 2: 167 observations,    complexity param=0.04613713
  mean=0.9130126, MSE=0.1562149
  left son=4 (40 obs) right son=5 (127 obs)
  Primary splits:
    BASECO3 splits as  LR-R-RR---RLLLLL-RLL-R-RLLRRRRRRR, improve=0.16778110, (0 missing)
    BASIN   splits as  LRRL-LR-RRLRRR, improve=0.09140733, (0 missing)
    ECO3    splits as  RLLRRL-RRR, improve=0.08472502, (0 missing)

Node number 3: 15 observations
  mean=1.892, MSE=3.706083

Node number 4: 40 observations
  mean=0.62454, MSE=0.01924522

Node number 5: 127 observations
  mean=1.00387, MSE=0.1648899
```

ANALYSIS: CHL-A SPEC VS. REGIONS (BASIN, ECOREGION III, BASIN-ECOREGION III)

```
Call:
mvpart(form = CHLASPEC ~ BASIN + ECO3 + BASECO3, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
 n=189 (575 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.30755466      0 1.0000000 1.0109540 0.10189580
2 0.03968847      1 0.6924453 0.8532553 0.08669967

Node number 1: 189 observations,    complexity param=0.3075547
  mean=15.71384, MSE=70.21389
  left son=2 (102 obs) right son=3 (87 obs)
  Primary splits:
      BASECO3 splits as  LR-L-L-RRRLLLLLLLLRLLLLRLLRRLRRRR, improve=0.3075547, (0 missing)
      BASIN   splits as  LRRLLLLRLRLRLR, improve=0.2557463, (0 missing)
      ECO3    splits as  LLRLRLRRRR, improve=0.0975637, (0 missing)

Node number 2: 102 observations
  mean=11.42211, MSE=27.40436

Node number 3: 87 observations
  mean=20.74552, MSE=73.49194
```

ANALYSIS: CHL-A FLUORO VS. REGIONS (BASIN, ECOREGION III, BASIN-ECOREGION III)

```
Call:
mvpart(form = CHLAFLUORO ~ BASIN + ECO3 + BASECO3, data = res,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
 n=108 (656 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.2220510      0  1.000000 1.010831 0.1921429
2 0.1157116      1  0.777949 1.318865 0.2527849

Node number 1: 108 observations,    complexity param=0.222051
  mean=16.56731, MSE=157.837
  left son=2 (89 obs) right son=3 (19 obs)
  Primary splits:
      BASECO3 splits as  LL-L-LLRR-L-LLL-R--RLL-RLLRRLRL-L, improve=0.2220510, (0 missing)
      BASIN   splits as  LLLL-LRLLRLRLL, improve=0.1927906, (0 missing)
      ECO3    splits as  L-RRRLLRRR, improve=0.1547249, (0 missing)
```

```
Node number 2: 89 observations
  mean=13.83197, MSE=107.1276

Node number 3: 19 observations
  mean=29.38026, MSE=196.1513
```

ANALYSIS: SECCHI TRANSPARENCY VS. REGIONS (BASIN, ECOREGION III, BASIN-ECOREGION III)

```
Call:
mvpart(form = SECCHI ~ BASIN + ECO3 + BASECO3, data = res, xval = 10,
    method = "anova", minsplit = 20, minbucket = 10)
  n=281 (483 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.2659223      0 1.0000000 1.0140252 0.1479417
2 0.0889435      1 0.7340777 0.8568287 0.1120344

Node number 1: 281 observations,    complexity param=0.2659223
  mean=0.8730194, MSE=0.2843646
  left son=2 (244 obs) right son=3 (37 obs)
  Primary splits:
      BASECO3 splits as  LLLL-LRLLLLLRRRRRLRLLRLLRLLLLLLLL, improve=0.2659223, (0 missing)
      BASIN   splits as  LLLR-RRLRLRLRL, improve=0.1687168, (0 missing)
      ECO3    splits as  RLLLLRLLLL, improve=0.1675910, (0 missing)

Node number 2: 244 observations
  mean=0.7659362, MSE=0.1577321

Node number 3: 37 observations
  mean=1.579189, MSE=0.5451602
```

Appendix 2.5. Cart, nonparametric changepoint, and linear regression analysis results for % difference in med/meansub and med/mean cen vs. Percent censored

TP Median % Difference vs. % Censored

```
CART
Call:
mvpart(form = PercDiff ~ PercCen, data = cp, xval = 10, method = "anova",
    minsplit = 10, minbucket = 5)
  n=171 (593 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.72416228      0 1.0000000 1.0053115 0.13532466
2 0.10427045      1 0.2758377 0.2882306 0.04819275
3 0.01596586      2 0.1715673 0.2001309 0.03500224

Node number 1: 171 observations,    complexity param=0.7241623
  mean=0.1372732, MSE=0.04644022
  left son=2 (140 obs) right son=3 (31 obs)
  Primary splits:
      PercCen < 0.5984941 to the left,  improve=0.7241623, (0 missing)

Node number 2: 140 observations,    complexity param=0.1042704
  mean=0.05097895, MSE=0.01082497
  left son=4 (112 obs) right son=5 (28 obs)
  Primary splits:
      PercCen < 0.4360338 to the left,  improve=0.5463825, (0 missing)

Node number 3: 31 observations
  mean=0.5269892, MSE=0.02177445

Node number 4: 112 observations
  mean=0.01252581, MSE=0.002179478

Node number 5: 28 observations
  mean=0.2047915, MSE=0.01583407

nCPA Threshold 1
cp      r2 mean left mean right pperm      5%       25%      50%      75%       95%
0.5985   0.7241542   0.05098  0.5269774 0.001   0.46725 0.57515 0.5952 0.5985 0.62455

nCPA Threshold 2
cp      r2  mean left mean right pperm      5%       25%      50%      75%           95%
0.4360338 0.5463825 0.01252581 0.2047915 0.001 0.3479313   0.4323708 0.4360338 0.4508065
0.4535714
```

Linear Regression Analysis

Nonlinear Regression **Tuesday, November 27, 2012, 1:59:49 PM**

Data Source: Fig1Reg Data in CenData_Figures
Equation: Polynomial, Linear
f = y0+a*x

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0.7987	0.6380	0.6327	13.9839

	Coefficient	Std. Error	t	P
y0	-46.9299	7.3552	-6.3805	<0.0001
a	1.3545	0.1228	11.0269	<0.0001

Analysis of Variance:

	DF	SS	MS
Regression	2	96859.3133	48429.6567
Residual	69	13492.8915	195.5492
Total	71	110352.2048	1554.2564

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	1	23777.2230	23777.2230	121.5921	<0.0001
Residual	69	13492.8915	195.5492		
Total	70	37270.1145	532.4302		

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.2981)

W Statistic= 0.9795 Significance Level = 0.0500

Constant Variance Test Passed (P = 0.0548)

TP Mean % Difference vs. %Censored

CART
Call:

```
mvpart(form = PercDiff ~ PercCen, data = tpchla2, xval = 10,
   method = "anova", minsplit = 10, minbucket = 5)
  n=171 (593 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.74946322      0 1.0000000 1.0107185 0.13214300
```

2 0.09878100 1 0.2505368 0.3139202 0.05115042
3 0.03563737 2 0.1517558 0.2100036 0.03785194

Node number 1: 171 observations, complexity param=0.7494632
mean=0.09494548, MSE=0.01957992
left son=2 (127 obs) right son=3 (44 obs)
Primary splits:
PercCen < 0.525 to the left, improve=0.7494632, (0 missing)

Node number 2: 127 observations
mean=0.02364292, MSE=0.002122357

Node number 3: 44 observations, complexity param=0.098781
mean=0.3007506, MSE=0.01293863
left son=6 (20 obs) right son=7 (24 obs)
Primary splits:
PercCen < 0.675 to the left, improve=0.5809512, (0 missing)

Node number 6: 20 observations
mean=0.2057766, MSE=0.004427554

Node number 7: 24 observations
mean=0.3798956, MSE=0.006250549

nCPA Threshold 1
cp r2 mean left mean right pperm 5% 25% 50% 75% 95%

0.5590278 0.7501962 0.02797440 0.3142757 0.001 0.4935897 0.5217593 0.5590278 0.575188 0.6245502

nCPA Threshold 2
cp r2 mean left mean right pperm 5% 25% 50% 75% 95%
0.6744422 0.5476624 0.2158460 0.3798956 0.001 0.6232436 0.6658215 0.6744422 0.7009085 0.7009085

Linear Regression Analysis

Nonlinear Regression **Tuesday, November 27, 2012, 2:01:27 PM**

Data Source: Fig1Reg Data in CenData_Figures
Equation: Polynomial, Linear
f = y0+a*x

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0.7881	0.6211	0.6128	7.3397

	Coefficient	Std. Error	t	P
y0	-36.4825	7.5939	-4.8042	<0.0001
a	0.9913	0.1142	8.6829	<0.0001

Analysis of Variance:

	DF	SS	MS
Regression	2	43902.2940	21951.1470
Residual	46	2478.0710	53.8711
Total	48	46380.3650	966.2576

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	1	4061.5212	4061.5212	75.3933	<0.0001
Residual	46	2478.0710	53.8711		
Total	47	6539.5922	139.1403		

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.0626)

W Statistic= 0.9549 Significance Level = 0.0500

Constant Variance Test Passed (P = 0.0961)

TP Standard Deviation %Difference vs. %Censored

nCPA Threshold 1

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%
0.4936	0.4176906	0.02962232	0.2403421	0.001	0.43225	0.4936	0.4936	0.56695

95%
0.5707

Chl-a Median %Difference vs. %Censored

CART

```
Call:
mvpart(form = CHLAPercDiff ~ CHLAPercCen, data = data1, xval = 10,
   method = "anova", minsplit = 10, minbucket = 5)
  n= 153

         CP nsplit rel error    xerror      xstd
1 0.5901786      0 1.0000000 1.0183400 0.2151634
2 0.1029272      1 0.4098214 0.4747669 0.1083238
```

Node number 1: 153 observations,   complexity param=0.5901786
  mean=0.07960784, MSE=0.02341553
  left son=2 (140 obs) right son=3 (13 obs)
  Primary splits:
      CHLAPercCen < 0.545 to the left, improve=0.5901786, (0 missing)

Node number 2: 140 observations
  mean=0.04378571, MSE=0.008444954

Node number 3: 13 observations
  mean=0.4653846, MSE=0.02199408

nCPA Threshold 1

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.5446429	0.5598393	0.04391427	0.4505385	0.001	0.3273810	0.5247059	0.5446429	0.5446429	0.5708333

nCPA Threshold 2

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.3273810	0.3061822	0.01846965	0.1456928	0.001	0.1550481	0.2321429	0.2970588	0.3273810	0.4258242

Linear Regression Analysis

Nonlinear Regression **Tuesday, November 27, 2012, 2:01:57 PM**

Data Source: Fig1Reg Data in CenData_Figures
Equation: Polynomial, Linear
f = y0+a*x

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0.5914	0.3498	0.3392	15.9185

	Coefficient	Std. Error	t	P
y0	-12.5082	5.6849	-2.2003	0.0316
a	0.7441	0.1299	5.7288	<0.0001

Analysis of Variance:

	DF	SS	MS
Regression	2	28650.0692	14325.0346
Residual	61	15457.3087	253.3985
Total	63	44107.3779	700.1171

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	1	8316.2956	8316.2956	32.8190	<0.0001

Residual	61	15457.3087	253.3985
Total	62	23773.6042	383.4452

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.1581)

W Statistic= 0.9719 Significance Level = 0.0500

Constant Variance Test Failed (P = 0.0018)

Chl-a Mean %Difference vs. %Censored

```
CART
Call:
mvpart(form = CHLAPercDiff ~ CHLAPerCen, data = chla2, xval = 10,
   method = "anova", minsplit = 10, minbucket = 5)
  n=154 (610 observations deleted due to missingness)

        CP nsplit rel error    xerror       xstd
1 0.6295481      0 1.0000000 1.0081680 0.18497884
2 0.1387389      1 0.3704519 0.5353974 0.11175735
3 0.0100000      2 0.2317130 0.3293759 0.07149875

Node number 1: 154 observations,    complexity param=0.6295481
  mean=0.05013182, MSE=0.01072591
  left son=2 (120 obs) right son=3 (34 obs)
  Primary splits:
      CHLAPerCen < 0.3825  to the left,  improve=0.6295481, (0 missing)

Node number 2: 120 observations
  mean=0.006391667, MSE=0.001249513

Node number 3: 34 observations,    complexity param=0.1387389
  mean=0.2045088, MSE=0.01358727
  left son=6 (20 obs) right son=7 (14 obs)
  Primary splits:
      CHLAPerCen < 0.54465 to the left,  improve=0.4960689, (0 missing)

Node number 6: 20 observations
  mean=0.13582, MSE=0.006624571

Node number 7: 14 observations
  mean=0.3026357, MSE=0.007164869
```

nCPA Threshold 1

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.3825	0.6295481	0.006391667	0.2045088	0.001	0.32735	0.3818	0.3825	0.54305	0.5625

nCPA Threshold 2

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.54465	0.4960689	0.13582	0.3026357	0.001	0.54305	0.54465	0.54465	0.5625	0.60435

Linear Regression Analysis

Nonlinear Regression **Monday, December 03, 2012, 11:16:40 AM**

Data Source: Fig1Reg Data in CenData_Figures
Equation: Polynomial, Linear
f = y0+a*x

R	**Rsqr**	**Adj Rsqr**	**Standard Error of Estimate**
0.6564	0.4308	0.4166	10.2247

	Coefficient	**Std. Error**	**t**	**P**
y0	-22.3800	7.3994	-3.0245	0.0043
a	0.7836	0.1424	5.5021	<0.0001

Confidence Intervals:

	Coefficient	**95% Conf-L**	**95% Conf-U**
y0	-22.3800	-37.3348	-7.4251
a	0.7836	0.4958	1.0715

Analysis of Variance:

	DF	**SS**	**MS**
Regression	2	15875.8655	7937.9328
Residual	40	4181.7992	104.5450
Total	42	20057.6647	477.5634

Corrected for the mean of the observations:

	DF	**SS**	**MS**	**F**	**P**
Regression	1	3164.9527	3164.9527	30.2736	<0.0001
Residual	40	4181.7992	104.5450		
Total	41	7346.7519	179.1891		

Statistical Tests:

Normality Test (Shapiro-Wilk) Passed (P = 0.8944)

W Statistic= 0.9865 Significance Level = 0.0500

Constant Variance Test Passed (P = 0.0818)

Chl-a Standard Deviation % Difference vs. %Censored

cp	r2	mean left	mean right	pperm	5%	25%	50%
0.38105	0.2934910	0.01450731	0.1746458	0.001	0.32735	0.3618	0.38105

75%	95%
0.4577	0.5247

APPENDIX 2.6. CART, nonparametric changepoint analysis results for biological response variables vs. nutrients

Dataset 1 Median Chl-a vs. TP

CART
```
Call:
mvpart(form = CHLA ~ TP, data = tpchla, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=162 (602 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.33505624      0 1.0000000 1.0121597 0.10719234
2 0.02665919      1 0.6649438 0.6887834 0.07998744

Node number 1: 162 observations,    complexity param=0.3350562
  mean=15.65565, MSE=70.42548
  left son=2 (99 obs) right son=3 (63 obs)
  Primary splits:
      TP < 0.063 to the left,  improve=0.3350562, (0 missing)

Node number 2: 99 observations
  mean=11.78061, MSE=32.69703

Node number 3: 63 observations
  mean=21.745, MSE=69.03633
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.063	0.3334566	11.79878	21.745	0.001	0.06175	0.063	0.063	0.0655	0.06733335

Dataset 1 Mean Chl-a vs. TP

CART
```
Call:
mvpart(form = CHLAMean ~ TPMean, data = tpchla, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 162

          CP nsplit rel error    xerror       xstd
1 0.32851571      0 1.0000000 1.0092339 0.09404065
2 0.02258872      1 0.6714843 0.6957194 0.07075345

Node number 1: 162 observations,    complexity param=0.3285157
  mean=18.65536, MSE=85.07234
```

```
 left son=2 (63 obs) right son=3 (99 obs)
 Primary splits:
    TPMean < 0.0625 to the left,  improve=0.3285157, (0 missing)

Node number 2: 63 observations
 mean=12.02832, MSE=26.9998

Node number 3: 99 observations
 mean=22.87257, MSE=76.29516
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.0625	0.3205938	12.06555	22.78234	0.001	0.0605	0.0615	0.0625	0.0625	0.0665

Dataset 2 Median Chl-a vs. TP

CART

```
Call:
mvpart(form = CHLA ~ TP, data = res1, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=184 (580 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.33611759      0 1.0000000 1.0119627 0.08750178
2 0.04452657      1 0.6638824 0.7566832 0.08203761

Node number 1: 184 observations,    complexity param=0.3361176
 mean=14.45764, MSE=90.9913
 left son=2 (85 obs) right son=3 (99 obs)
 Primary splits:
    TP < 0.039 to the left,  improve=0.3361176, (0 missing)

Node number 2: 85 observations
 mean=8.489294, MSE=31.48971

Node number 3: 99 observations
 mean=19.58197, MSE=85.23595
```

nCPA

	cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
[1,]	0.039	0.3361176	8.489294	19.58197	0.001	0.0325	0.039	0.0555	0.063	0.06683335

Dataset 2 Mean Chl-a vs. TP

```
CART
Call:
mvpart(form = CHLA ~ TP, data = res2, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=184 (580 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.33769484      0 1.0000000 1.0216041 0.08414968
2 0.02335195      1 0.6623052 0.7783673 0.07708919

Node number 1: 184 observations,    complexity param=0.3376948
  mean=17.38129, MSE=94.74559
  left son=2 (78 obs) right son=3 (106 obs)
  Primary splits:
      TP < 0.05303846 to the left,  improve=0.3376948, (0 missing)

Node number 2: 78 observations
  mean=10.78731, MSE=34.66139

Node number 3: 106 observations
  mean=22.23346, MSE=83.41983

nCPA
             cp        r2 mean left mean right pperm         5%        25%        50%        75%        95%

[1,] 0.05303846 0.3376948  10.78731   22.23346 0.001 0.04526356 0.05012108 0.05303846 0.05367704
0.06770609
```

Dataset 3 Median Chl-a vs. TP

```
CART
Call:
mvpart(form = CHLA ~ TP, data = res3, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=113 (651 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.28035509      0 1.0000000 1.0195579 0.11127591
2 0.04215627      1 0.7196449 0.7542813 0.08775915

Node number 1: 113 observations,    complexity param=0.2803551
  mean=20.2731, MSE=87.98518
  left son=2 (19 obs) right son=3 (94 obs)
  Primary splits:
      TP < 0.05975 to the left,  improve=0.2803551, (0 missing)
```

```
Node number 2: 19 observations
  mean=9.226053, MSE=24.90836

Node number 3: 94 observations
  mean=22.50601, MSE=71.08176
```

nCPA

```
        cp       r2 mean left mean right pperm    5%    25%     50%     75%    95%

[1,] 0.05975 0.2803551  9.226053   22.50601 0.001 0.048 0.0585 0.05975 0.05975 0.0653
```

Dataset 3 Mean Chl-a vs. TP

CART

```
Call:
mvpart(form = CHLA ~ TP, data = res4, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=113 (651 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.21683832      0 1.0000000 1.0148421 0.1060045
2 0.04304475      1 0.7831617 0.8910112 0.1000873

Node number 1: 113 observations,    complexity param=0.2168383
  mean=23.31431, MSE=102.4765
  left son=2 (14 obs) right son=3 (99 obs)
  Primary splits:
      TP < 0.06008769 to the left,  improve=0.2168383, (0 missing)

Node number 2: 14 observations
  mean=10.77904, MSE=31.0429

Node number 3: 99 observations
  mean=25.08697, MSE=87.21506
```

nCPA

```
          cp       r2 mean left mean right pperm       5%       25%       50%       75%       95%

[1,] 0.06008769 0.2168383  10.77904   25.08697 0.001 0.05428571 0.06008769 0.06464293 0.07134091
0.07830923
```

Dataset 4 Median Chl-a vs. TP

```
CART
Call:
mvpart(form = CHLACen ~ TPCen, data = tpchla, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=124 (640 observations deleted due to missingness)

         CP nsplit rel error   xerror     xstd
1 0.27266242      0 1.0000000 1.012375 0.104848
2 0.03530451      1 0.7273376 0.833448 0.101186

Node number 1: 124 observations,    complexity param=0.2726624
 mean=16.77309, MSE=81.92725
 left son=2 (66 obs) right son=3 (58 obs)
 Primary splits:
    TPCen < 0.06275 to the left,  improve=0.2726624, (0 missing)

Node number 2: 66 observations
 mean=12.34242, MSE=45.96712

Node number 3: 58 observations
 mean=21.81488, MSE=75.08927

nCPA
cp       r2 mean left mean right   pperm    5%      25%      50%      75%      95%

0.06275 0.2726624  12.34242   21.81488 0.001  0.048    0.053    0.0615  0.06275 0.068
```

Dataset 4 Mean Chl-a vs. TP

```
CART
Call:
mvpart(form = CHLAMeanCen ~ TPMeanCen, data = tpchla, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
 n=125 (37 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.21885060      0 1.0000000 1.0131612 0.09503229
2 0.03893098      1 0.7811494 0.8497047 0.08744799

Node number 1: 125 observations,    complexity param=0.2188506
 mean=20.34432, MSE=92.30988
 left son=2 (28 obs) right son=3 (97 obs)
 Primary splits:
    TPMeanCen < 0.0535 to the left,  improve=0.2188506, (0 missing)
```

```
Node number 2: 28 observations
  mean=11.97857, MSE=30.40978

Node number 3: 97 observations
  mean=22.75918, MSE=84.14435
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.0535	0.2188506	11.97857	22.75918	0.001	0.0535	0.0535	0.0555	0.0655	0.0835

Dataset 5 Median Chl-a vs. TP

CART

```
mvpart(form = CHLACen2 ~ TPCen2, data = tpchla, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=162 (602 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.32490082      0 1.0000000 1.0241321 0.10108218
2 0.04701995      1 0.6750992 0.8393252 0.09352402
3 0.02474351      2 0.6280792 0.7673109 0.09058549

Node number 1: 162 observations,    complexity param=0.3249008
  mean=14.46795, MSE=83.66457
  left son=2 (76 obs) right son=3 (86 obs)
  Primary splits:
      TPCen2 < 0.0485   to the left,  improve=0.3249008, (0 missing)

Node number 2: 76 observations,    complexity param=0.02474351
  mean=8.921839, MSE=27.8853
  left son=4 (30 obs) right son=5 (46 obs)
  Primary splits:
      TPCen2 < 0.017104 to the left,  improve=0.1582446, (0 missing)

Node number 3: 86 observations,    complexity param=0.04701995
  mean=19.36917, MSE=81.75327
  left son=6 (26 obs) right son=7 (60 obs)
  Primary splits:
      TPCen2 < 0.06275  to the left,  improve=0.09064319, (0 missing)

Node number 4: 30 observations
  mean=6.320659, MSE=16.43765

Node number 5: 46 observations
```

```
  mean=10.61826, MSE=28.06062

Node number 6: 26 observations
  mean=15.23385, MSE=49.42196

Node number 7: 60 observations
  mean=21.16114, MSE=85.14197
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.0485	0.3322535	8.92184	19.4351	0.001	0.039	0.0485	0.053	0.06275	0.06425

Datatset 5 Mean Chl-a vs. TP

CART

```
Call:
mvpart(form = CHLAMeanCen2 ~ TPMeanCen2, data = tpchla, xval = 10,
    method = "anova", minsplit = 20, minbucket = 10)
  n= 162

          CP nsplit rel error    xerror       xstd
1 0.34673261      0 1.0000000 1.0147464 0.08854691
2 0.03377178      1 0.6532674 0.6920097 0.07013563

Node number 1: 162 observations,    complexity param=0.3467326
  mean=17.87212, MSE=98.29223
  left son=2 (64 obs) right son=3 (98 obs)
  Primary splits:
      TPMeanCen2 < 0.0535 to the left,  improve=0.3467326, (0 missing)

Node number 2: 64 observations
  mean=10.64808, MSE=30.745

Node number 3: 98 observations
  mean=22.58986, MSE=86.06653
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.0535	0.3467326	10.64808	22.58986	0.001	0.051975	0.0535	0.054	0.0555	0.065525

Dataset 1 Median Secchi depth vs. TP

```
CART
Call:
mvpart(form = SECCHI ~ TP, data = secchi, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 203

          CP nsplit rel error    xerror      xstd
1 0.27115484      0 1.0000000 1.0152160 0.1586967
2 0.03029607      1 0.7288452 0.7390394 0.1265506
Node number 1: 203 observations,    complexity param=0.2711548
  mean=0.9419192, MSE=0.3356777
  left son=2 (74 obs) right son=3 (129 obs)
  Primary splits:
      TP < 0.06125 to the right, improve=0.2711548, (0 missing)

Node number 2: 74 observations
  mean=0.5435838, MSE=0.03802539

Node number 3: 129 observations,    complexity param=0.03029607
  mean=1.170422, MSE=0.36319
  left son=6 (93 obs) right son=7 (36 obs)
  Primary splits:
      TP < 0.056   to the right, improve=0.04406374, (0 missing)

Node number 6: 93 observations
  mean=1.091714, MSE=0.3689875

Node number 7: 36 observations
  mean=1.37375, MSE=0.2908672

nCPA
cp       r2 mean left mean right   pperm      5%      25%      50%      75%       95%
0.06125 0.2711548  1.170422  0.5435838 0.001   0.06125 0.06125 0.06125 0.0625 0.06425
```

Dataset 1 Mean Secchi depth vs. TP

```
CART
Call:
mvpart(form = SECCHI ~ TP, data = means, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=220 (544 observations deleted due to missingness)
```

```
        CP nsplit rel error  xerror      xstd
1 0.31983505      0 1.0000000 1.00690 0.1777562
2 0.03393441      1 0.6801649 0.78548 0.1818678

Node number 1: 220 observations,    complexity param=0.3198351
  mean=0.9635636, MSE=0.3550924
  left son=2 (129 obs) right son=3 (91 obs)
  Primary splits:
      TP < 0.06363456 to the right, improve=0.3198351, (0 missing)

Node number 2: 129 observations
  mean=0.6805159, MSE=0.102488

Node number 3: 91 observations
  mean=1.364807, MSE=0.4386127
```

nCPA

```
            cp        r2 mean left mean right pperm        5%        25%
[1,] 0.06363456 0.3360165  1.379971  0.6805159 0.001 0.06024821 0.06359886
            50%        75%        95%
[1,] 0.06363456 0.06383091 0.0650508
```

Dataset 2 Median Secchi depth vs. TP

CART
Call:

```
mvpart(form = SECCHI ~ TP, data = res1, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=218 (546 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.35566840      0 1.0000000 1.005318 0.1538925
2 0.03417263      1 0.6443316 0.655134 0.1079087

Node number 1: 218 observations,    complexity param=0.3556684
  mean=0.9252734, MSE=0.3269468
  left son=2 (118 obs) right son=3 (100 obs)
  Primary splits:
      TP < 0.039 to the right, improve=0.3556684, (0 missing)

Node number 2: 118 observations
  mean=0.6113525, MSE=0.05320207

Node number 3: 100 observations
```

```
  mean=1.2957, MSE=0.396465

nCPA
       cp       r2 mean left mean right pperm    5%   25%   50%   75%   95%

[1,] 0.039 0.3556684    1.2957  0.6113525 0.001 0.038 0.039 0.039 0.045 0.048
```

Dataset 2 Mean Secchi depth vs. TP

```
CART
Call:
mvpart(form = SECCHI ~ TP, data = res2, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=218 (546 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.37383435      0 1.0000000 1.0105385 0.1792634
2 0.07121586      1 0.6261656 0.7503075 0.1334056

Node number 1: 218 observations,    complexity param=0.3738344
  mean=0.9688709, MSE=0.3529719
  left son=2 (141 obs) right son=3 (77 obs)
  Primary splits:
      TP < 0.04488776 to the right, improve=0.3738344, (0 missing)

Node number 2: 141 observations
  mean=0.700432, MSE=0.09930532

Node number 3: 77 observations
  mean=1.460428, MSE=0.4438969

nCPA

           cp       r2 mean left mean right pperm       5%       25%       50%       75%       95%

[1,] 0.04488776 0.3738344  1.460428   0.700432 0.001 0.02926875 0.0414377 0.04361905 0.04488776
0.04616528
```

Dataset 3 Median Secchi depth vs. TP

```
CART
mvpart(form = SECCHI ~ TP, data = res3, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=155 (609 observations deleted due to missingness)
```

```
        CP nsplit rel error    xerror      xstd
1 0.2232995      0 1.0000000 1.0127410 0.1799265
2 0.1082444      1 0.7767005 0.9207133 0.1694218

Node number 1: 155 observations,    complexity param=0.2232995
  mean=0.7649652, MSE=0.1976994
  left son=2 (134 obs) right son=3 (21 obs)
  Primary splits:
      TP < 0.0535 to the right, improve=0.2232995, (0 missing)

Node number 2: 134 observations,    complexity param=0.1082444
  mean=0.6817881, MSE=0.134388
  left son=4 (29 obs) right son=5 (105 obs)
  Primary splits:
      TP < 0.142  to the right, improve=0.1841947, (0 missing)

Node number 3: 21 observations
  mean=1.295714, MSE=0.2758459

Node number 4: 29 observations
  mean=0.3824138, MSE=0.03194762

Node number 5: 105 observations
  mean=0.7644724, MSE=0.1310908
```

nCPA

```
        cp       r2 mean left mean right pperm    5%   25%    50%   75%   95%

[1,] 0.0535 0.2232995  1.295714   0.681788 0.001 0.043 0.043 0.0535 0.055 0.145
```

Dataset 3 Mean Secchi depth vs. TP

CART

```
Call:
mvpart(form = SECCHI ~ TP, data = res4, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=155 (609 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.19146257      0 1.0000000 1.011167 0.1762489
2 0.03575207      1 0.8085374 0.882221 0.1509757

Node number 1: 155 observations,    complexity param=0.1914626
  mean=0.8027671, MSE=0.1977741
  left son=2 (133 obs) right son=3 (22 obs)
```

```
  Primary splits:
     TP < 0.06454724 to the right, improve=0.1914626, (0 missing)

Node number 2: 133 observations
  mean=0.7236241, MSE=0.1413403

Node number 3: 22 observations
  mean=1.281222, MSE=0.2721564
```

nCPA

```
          cp        r2 mean left mean right pperm       5%        25%        50%        75%        95%
[1,] 0.06454724 0.1914626  1.281222  0.7236241 0.001 0.050875 0.06017343 0.06454724 0.06534965
0.09960142
```

Dataset 4 Median Secchi depth vs. TP

CART

```
Call:
mvpart(form = SECCHI ~ TPCEN, data = secchi, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=161 (42 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.30833813      0 1.0000000 1.016640 0.1890211
2 0.07944764      1 0.6916619 0.776002 0.1269903

Node number 1: 161 observations,    complexity param=0.3083381
  mean=0.7680099, MSE=0.1771376
  left son=2 (106 obs) right son=3 (55 obs)
  Primary splits:
     TPCEN < 0.0485 to the right, improve=0.3083381, (0 missing)

Node number 2: 106 observations
  mean=0.599666, MSE=0.045736

Node number 3: 55 observations
  mean=1.092455, MSE=0.2705017
```

nCPA

```
cp    r2   mean left   mean right    pperm     5%        25%        50%        75%        95%
```

```
0.0485  0.3083381  1.092455   0.599666 0.001      0.0215   0.0385    0.0445    0.0485    0.0485
```

Dataset 4 Mean Secchi depth vs. TP

```
Call:
mvpart(form = SECCHI ~ TPDEL, data = means, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=128 (636 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.2318121      0 1.0000000 1.0096938 0.2552929
2 0.1063129      1 0.7681879 0.9883239 0.2623325

Node number 1: 128 observations,    complexity param=0.2318121
  mean=0.6950821, MSE=0.117738
  left son=2 (108 obs) right son=3 (20 obs)
  Primary splits:
      TPDEL < 0.06280089 to the right, improve=0.2318121, (0 missing)

Node number 2: 108 observations
  mean=0.6239886, MSE=0.06420295

Node number 3: 20 observations
  mean=1.078987, MSE=0.2321512


nCPA
         cp       r2 mean left mean right pperm     5%    25%   50%    75%
[1,] 0.0475 0.2528475  1.235897  0.7053345 0.001 0.0415 0.0475 0.048 0.0505
       95%
[1,] 0.097
```

Dataset 5 Median Secchi depth vs. TP

```
Call:
mvpart(form = SECCHI ~ TPCORR, data = secchi, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 203

          CP nsplit rel error    xerror      xstd
1 0.43765497      0  1.000000 1.0064521 0.1584836
2 0.05995859      1  0.562345 0.6036443 0.1004478

Node number 1: 203 observations,    complexity param=0.437655
  mean=0.9419192, MSE=0.3356777
  left son=2 (146 obs) right son=3 (57 obs)
  Primary splits:
```

```
    TPCORR < 0.025184 to the right, improve=0.437655, (0 missing)

Node number 2: 146 observations,    complexity param=0.05995859
  mean=0.7024288, MSE=0.1122438
  left son=4 (106 obs) right son=5 (40 obs)
  Primary splits:
      TPCORR < 0.0485   to the right, improve=0.2493186, (0 missing)

Node number 3: 57 observations
  mean=1.555351, MSE=0.3847727

Node number 4: 106 observations
  mean=0.599666, MSE=0.045736

Node number 5: 40 observations
  mean=0.97475, MSE=0.1863462
```

nCPA

cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
0.025184	0.437655	1.555351	0.7024288	0.001	0.021	0.0215	0.025184	0.0385	0.047

Dataset 5 Mean Secchi depth vs. TP

CART

```
Call:
mvpart(form = SECCHI ~ TPCEN, data = means, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=166 (598 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.25255277      0 1.0000000 1.0096439 0.1800954
2 0.07561781      1 0.7474472 0.7836953 0.1220721

Node number 1: 166 observations,    complexity param=0.2525528
  mean=0.8033626, MSE=0.1769025
  left son=2 (134 obs) right son=3 (32 obs)
  Primary splits:
      TPCEN < 0.0475 to the right, improve=0.2525528, (0 missing)

Node number 2: 134 observations
  mean=0.7000708, MSE=0.09587795

Node number 3: 32 observations
  mean=1.235897, MSE=0.2844299
```

```
nCPA
[1,] 0.04198334 0.3750554  1.519185  0.7380332 0.001 0.03943326 0.04194167
           50%    75%    95%
[1,] 0.04198334 0.0475 0.0485
```

Dataset 1 Median Chl-a vs. TN

```
CART
Call:
mvpart(form = CHLA ~ TN, data = tn1, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 133

          CP nsplit rel error    xerror       xstd
1 0.50269326      0 1.0000000 1.0161060 0.11602497
2 0.05319156      1 0.4973067 0.5450037 0.06967514

Node number 1: 133 observations,    complexity param=0.5026933
  mean=15.70831, MSE=74.99601
  left son=2 (72 obs) right son=3 (61 obs)
  Primary splits:
      TN < 0.90125 to the left,  improve=0.5026933, (0 missing)

Node number 2: 72 observations
  mean=10.05674, MSE=14.88943

Node number 3: 61 observations
  mean=22.37902, MSE=63.74314

nCPA
          cp        r2 mean left mean right pperm  5%     25%    50%   75%   95%
[1,] 0.90125 0.5026933  10.05674   22.37902 0.001 0.8 0.90125 0.9025 0.967 1.015
```

Dataset 1 Mean Chl-a vs. TN

```
CART
mvpart(form = CHLA ~ TN, data = means1, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 137

          CP nsplit rel error    xerror       xstd
1 0.47913517      0 1.0000000 1.0103379 0.09451670
2 0.04159491      1 0.5208648 0.6438347 0.06688406
3 0.03589600      2 0.4792699 0.5632177 0.06194796
```

```
Node number 1: 137 observations,    complexity param=0.4791352
  mean=18.94164, MSE=93.68756
  left son=2 (63 obs) right son=3 (74 obs)
  Primary splits:
      TN < 0.8816964 to the left,  improve=0.4791352, (0 missing)

Node number 2: 63 observations
  mean=11.68032, MSE=22.35366

Node number 3: 74 observations,    complexity param=0.04159491
  mean=25.12357, MSE=71.31246
  left son=6 (15 obs) right son=7 (59 obs)
  Primary splits:
      TN < 0.9902598 to the left,  improve=0.1011686, (0 missing)

Node number 6: 15 observations
  mean=19.79653, MSE=27.07639

Node number 7: 59 observations
  mean=26.47791, MSE=73.51012
```

nCPA

```
            cp        r2 mean left mean right pperm       5%      25%       50%       75%
[1,] 0.8816964 0.4791352  11.68032   25.12357 0.001 0.8149765 0.880125 0.8817194 0.9583376
           95%
[1,] 0.9863913
```

Dataset 2 Median Chl-a vs. TN

CART

```
Call:
mvpart(form = CHLA ~ TN, data = half, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=129 (635 observations deleted due to missingness)

          CP nsplit rel error    xerror       xstd
1 0.53051587      0 1.0000000 1.0136162 0.09965980
2 0.04302063      1 0.4694841 0.5250788 0.07041855

Node number 1: 129 observations,    complexity param=0.5305159
  mean=14.42717, MSE=98.03679
  left son=2 (70 obs) right son=3 (59 obs)
  Primary splits:
      TN < 0.89575 to the left,  improve=0.5305159, (0 missing)
```

```
Node number 2: 70 observations
  mean=7.806214, MSE=22.42835

Node number 3: 59 observations
  mean=22.28254, MSE=74.02478

nCPA
         cp       r2 mean left mean right pperm     5%     25%     50%     75%    95%

[1,] 0.89575 0.5305159  7.806214   22.28254 0.001 0.7875 0.88375 0.89575 0.94375 0.9455
```

Dataset 2 Mean Chl-a vs. TN

```
CART
Call:
mvpart(form = CHLA ~ TN, data = meanshalf, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=129 (635 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.48270780      0 1.0000000 1.0134663 0.09042664
2 0.03122881      1 0.5172922 0.5954956 0.07059533

Node number 1: 129 observations,    complexity param=0.4827078
  mean=17.87512, MSE=106.6545
  left son=2 (59 obs) right son=3 (70 obs)
  Primary splits:
      TN < 0.869375 to the left,  improve=0.4827078, (0 missing)

Node number 2: 59 observations
  mean=10.05966, MSE=26.15421

Node number 3: 70 observations
  mean=24.46244, MSE=79.62898

nCPA
        cp       r2 mean left mean right pperm       5%      25%      50%       75%       95%
[1,] 0.869375 0.4827078  10.05966   24.46244 0.001 0.8337011 0.836962 0.869375 0.9381033
0.9509293
```

Dataset 3 Median Chl-a vs. TN

```
CART
mvpart(form = CHLA ~ TN, data = del, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=59 (705 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.55966065      0 1.0000000 1.0613225 0.1467142
2 0.05306205      1 0.4403394 0.5817974 0.1134840

Node number 1: 59 observations,    complexity param=0.5596606
  mean=18.05542, MSE=108.0551
  left son=2 (22 obs) right son=3 (37 obs)
  Primary splits:
      TN < 0.7875 to the left,  improve=0.5596606, (0 missing)

Node number 2: 22 observations
  mean=7.970455, MSE=27.63043

Node number 3: 37 observations
  mean=24.05189, MSE=59.44337

nCPA
         cp        r2 mean left mean right pperm    5%  25%    50%    75%    95%

[1,] 0.7875 0.5596606  7.970455   24.05189 0.001 0.715 0.73 0.7875 0.7875 0.9625
```

Dataset 3 Mean Chl-a vs. TN

```
CART
Call:
mvpart(form = CHLA ~ TN, data = meansdel, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=59 (705 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.53766364      0 1.0000000 1.0236522 0.1493543
2 0.05612623      1 0.4623364 0.6130226 0.1079974

Node number 1: 59 observations,    complexity param=0.5376636
  mean=20.82253, MSE=115.229
  left son=2 (21 obs) right son=3 (38 obs)
  Primary splits:
      TN < 0.8033301 to the left,  improve=0.5376636, (0 missing)

Node number 2: 21 observations
```

```
 mean=10.23442, MSE=34.48085

Node number 3: 38 observations
 mean=26.67385, MSE=63.66058
```

nCPA

```
        cp        r2 mean left mean right pperm       5%       25%       50%       75%       95%
[1,] 0.8033301 0.5376636  10.23442   26.67385 0.001 0.6990094 0.7067053 0.8033301 0.8366112
0.9777894
```

Dataset 4 Median Chl-a vs. TN

CART

```
Call:
mvpart(form = CHLACEN ~ TN, data = tn3, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n= 122

        CP nsplit rel error    xerror      xstd
1 0.48836779      0 1.0000000 1.0101653 0.10888941
2 0.04174536      1 0.5116322 0.5495606 0.07717422

Node number 1: 122 observations,    complexity param=0.4883678
 mean=15.27307, MSE=87.99076
 left son=2 (62 obs) right son=3 (60 obs)
 Primary splits:
    TN < 0.90125 to the left,  improve=0.4883678, (0 missing)

Node number 2: 62 observations
 mean=8.824371, MSE=21.53

Node number 3: 60 observations
 mean=21.93672, MSE=69.29077
```

nCPA

```
       cp       r2 mean left mean right pperm     5%     25%     50%    75%     95%
[1,] 0.90125 0.4883678  8.824371   21.93672 0.001 0.7925 0.89875 0.90125 0.9375 0.97025
```

Dataset 4 Mean Chl-a vs. TN

CART

```
mvpart(form = CHLACEN ~ TN, data = means3, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n= 124
```

```
         CP nsplit rel error    xerror       xstd
1 0.45263698      0 1.0000000 1.0248193 0.09483966
2 0.03489611      1 0.5473630 0.6683193 0.07623506
3 0.03253772      3 0.4775708 0.6803378 0.08349817
4 0.01000000      4 0.4450331 0.5696325 0.07617812

Node number 1: 124 observations,    complexity param=0.452637
  mean=18.92677, MSE=103.3332
  left son=2 (53 obs) right son=3 (71 obs)
  Primary splits:
      TN < 0.8816964 to the left,  improve=0.452637, (0 missing)

Node number 2: 53 observations,    complexity param=0.03253772
  mean=11.01113, MSE=26.68025
  left son=4 (31 obs) right son=5 (22 obs)
  Primary splits:
      TN < 0.6598083 to the left,  improve=0.2948375, (0 missing)

Node number 3: 71 observations,    complexity param=0.03489611
  mean=24.83563, MSE=78.86593
  left son=6 (31 obs) right son=7 (40 obs)
  Primary splits:
      TN < 1.057525  to the left,  improve=0.07666033, (0 missing)

Node number 4: 31 observations
  mean=8.648387, MSE=11.84428

Node number 5: 22 observations
  mean=14.34045, MSE=28.63475

Node number 6: 31 observations
  mean=22.04258, MSE=37.19375

Node number 7: 40 observations,    complexity param=0.03489611
  mean=27.00025, MSE=100.4304
  left son=14 (16 obs) right son=15 (24 obs)
  Primary splits:
      TN < 1.227029  to the right, improve=0.1157546, (0 missing)

Node number 14: 16 observations
  mean=22.82438, MSE=107.5304

Node number 15: 24 observations
  mean=29.78417, MSE=76.3216
```

```
nCPA
         cp      r2 mean left mean right pperm       5%      25%       50%       75%
[1,] 0.8816964 0.452637  11.01113   24.83563 0.001 0.8095752 0.837266 0.8816964 0.9391391
         95%
[1,] 0.9804882
```

Dataset 5 Median Chl-a vs. TN

```
CART
Call:
mvpart(form = CHLACORR ~ TN, data = tn4, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 133

         CP nsplit rel error   xerror       xstd
1 0.50023537      0 1.0000000 1.009647 0.10674569
2 0.04301734      1 0.4997646 0.525607 0.07223928

Node number 1: 133 observations,    complexity param=0.5002354
  mean=14.36413, MSE=89.88857
  left son=2 (72 obs) right son=3 (61 obs)
  Primary splits:
      TN < 0.90125 to the left,  improve=0.5002354, (0 missing)

Node number 2: 72 observations
  mean=8.191955, MSE=21.0419

Node number 3: 61 observations
  mean=21.64932, MSE=73.11081

nCPA
        cp       r2 mean left mean right pperm     5%    25%     50%    75%     95%
[1,] 0.90125 0.5002354  8.191955   21.64932 0.001 0.7925 0.8095 0.90125 0.9025 0.97025
```

Dataset 5 Mean Chl-a vs. TN

```
CART
Call:
mvpart(form = CHLACORR ~ TN, data = means4, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n= 135

         CP nsplit rel error    xerror       xstd
```

```
1 0.47795363      0 1.0000000 1.0211734 0.09069436
2 0.03302183      1 0.5220464 0.5716307 0.06623577

Node number 1: 135 observations,    complexity param=0.4779536
  mean=17.87338, MSE=107.4333
  left son=2 (63 obs) right son=3 (72 obs)
  Primary splits:
      TN < 0.8816964 to the left,  improve=0.4779536, (0 missing)

Node number 2: 63 observations
  mean=10.21286, MSE=25.84618

Node number 3: 72 observations
  mean=24.57634, MSE=82.54429
```

nCPA

```
          cp        r2 mean left mean right pperm       5%      25%       50%      75%
[1,] 0.8816964 0.4779536  10.21286   24.57634 0.001 0.8093527 0.837266 0.8816964 0.886534
         95%
[1,] 0.9623597
```

Dataset 1 Median Secchi depth vs. TN

CART

```
Call:
mvpart(form = SECCHI ~ TN, data = med, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=181 (583 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.31884118      0 1.0000000 1.010594 0.1625525
2 0.08511968      1 0.6811588 0.767809 0.1236868

Node number 1: 181 observations,    complexity param=0.3188412
  mean=0.966158, MSE=0.3546729
  left son=2 (140 obs) right son=3 (41 obs)
  Primary splits:
      TN < 0.60075 to the right, improve=0.3188412, (0 missing)

Node number 2: 140 observations
  mean=0.7841757, MSE=0.1960691

Node number 3: 41 observations
  mean=1.587561, MSE=0.3970209
```

```
nCPA
      cp       r2 mean left mean right pperm      5%     25%  50%     75%
[1,] 0.60075 0.3188412  1.587561  0.7841757 0.001 0.54475 0.60075 0.61 0.77035
     95%
[1,] 0.831
```

Dataset 1 Mean Secchi depth vs. TN

```
CART
Call:
mvpart(form = SECCHI ~ TN, data = means, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=181 (583 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.26400191      0 1.0000000 1.0122193 0.1897069
2 0.08719208      1 0.7359981 0.8509752 0.1665592

Node number 1: 181 observations,    complexity param=0.2640019
  mean=1.008466, MSE=0.3875842
  left son=2 (117 obs) right son=3 (64 obs)
  Primary splits:
      TN < 0.8072141 to the right, improve=0.2640019, (0 missing)

Node number 2: 117 observations
  mean=0.771883, MSE=0.1906745

Node number 3: 64 observations
  mean=1.440969, MSE=0.4581775

nCPA
         cp       r2 mean left mean right pperm       5%       25%       50%
[1,] 0.807214 0.2640019  1.440969   0.771883 0.001 0.562157 0.6146331 0.8027643
         75%       95%
[1,] 0.8083053 0.8895536
```

Dataset 2 Median Secchi depth vs. TN

```
CART
Call:
mvpart(form = SECCHI ~ TN, data = half, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=169 (595 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
```

```
1 0.34311972      0 1.0000000 1.0035719 0.16246240
2 0.07772987      1 0.6568803 0.8062548 0.12261249
3 0.06879287      2 0.5791504 0.7357520 0.12012580
4 0.01111935      3 0.5103575 0.5909253 0.09441448

Node number 1: 169 observations,    complexity param=0.3431197
  mean=0.9856781, MSE=0.3671651
  left son=2 (135 obs) right son=3 (34 obs)
  Primary splits:
      TN < 0.580625 to the right, improve=0.3431197, (0 missing)

Node number 2: 135 observations,    complexity param=0.06879287
  mean=0.8075526, MSE=0.2048831
  left son=4 (96 obs) right son=5 (39 obs)
  Primary splits:
      TN < 0.8185   to the right, improve=0.1543305, (0 missing)

Node number 3: 34 observations,    complexity param=0.07772987
  mean=1.692941, MSE=0.3853178
  left son=6 (22 obs) right son=7 (12 obs)
  Primary splits:
      TN < 0.4865   to the right, improve=0.3681613, (0 missing)

Node number 4: 96 observations
  mean=0.6942146, MSE=0.1334119

Node number 5: 39 observations
  mean=1.086538, MSE=0.2713592

Node number 6: 22 observations
  mean=1.414773, MSE=0.169117

Node number 7: 12 observations
  mean=2.202917, MSE=0.3797519

nCPA
          cp        r2 mean left mean right pperm      5%     25%   50%   75%     95%
[1,] 0.580625 0.3431197  1.692941  0.8075526 0.001 0.48125 0.51125 0.585 0.615 0.82075
```

Dataset 2 Mean Secchi transparency vs. TN

```
CART
Call:
mvpart(form = SECCHI ~ TN, data = meanshalf, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
```

```
  n=169 (595 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.3009112      0 1.0000000 1.0092939 0.1921936
2 0.1023279      1 0.6990888 0.8326075 0.1590778

Node number 1: 169 observations,    complexity param=0.3009112
  mean=1.028846, MSE=0.3991846
  left son=2 (153 obs) right son=3 (16 obs)
  Primary splits:
      TN < 0.5572045 to the right, improve=0.3009112, (0 missing)

Node number 2: 153 observations
  mean=0.916768, MSE=0.2526626

Node number 3: 16 observations
  mean=2.100592, MSE=0.5315432

nCPA
            cp        r2 mean left mean right pperm       5%       25%       50%       75%       95%
[1,] 0.5572045 0.3009112  2.100592   0.916768 0.001 0.5307276 0.547625 0.5750716 0.7897508
0.8713036
```

Dataset 3 Median Secchi transparency vs. TN

```
CART
Call:
mvpart(form = SECCHI ~ TN, data = del, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=71 (693 observations deleted due to missingness)

         CP nsplit rel error    xerror       xstd
1 0.5813137      0 1.0000000 1.0103899 0.22803634
2 0.1358861      1 0.4186863 0.6283537 0.16055040
3 0.0250810      2 0.2828003 0.3911901 0.09386974

Node number 1: 71 observations,    complexity param=0.5813137
  mean=1.088023, MSE=0.5249474
  left son=2 (59 obs) right son=3 (12 obs)
  Primary splits:
      TN < 0.58075 to the right, improve=0.5813137, (0 missing)

Node number 2: 59 observations,    complexity param=0.1358861
  mean=0.8388915, MSE=0.2020889
  left son=4 (42 obs) right son=5 (17 obs)
```

```
 Primary splits:
    TN < 0.795   to the right, improve=0.4247707, (0 missing)

Node number 3: 12 observations
 mean=2.312917, MSE=0.3068102

Node number 4: 42 observations
 mean=0.6524905, MSE=0.05859421

Node number 5: 17 observations
 mean=1.299412, MSE=0.2586849
```

nCPA

```
        cp       r2 mean left mean right pperm    5%     25%    50%   75%    95%
[1,] 0.58075 0.5813137  2.312917  0.8388915 0.001 0.575 0.58075 0.5825 0.795 0.8175
```

Dataset 3 Mean Secchi transparency vs. TN

CART

```
Call:
mvpart(form = SECCHI ~ TN, data = meansdel, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=71 (693 observations deleted due to missingness)

        CP nsplit rel error    xerror      xstd
1 0.47286093      0 1.0000000 1.0255171 0.2687084
2 0.08449067      1 0.5271391 0.6686414 0.1975607

Node number 1: 71 observations,    complexity param=0.4728609
 mean=1.132845, MSE=0.6004748
 left son=2 (56 obs) right son=3 (15 obs)
 Primary splits:
    TN < 0.6420038 to the right, improve=0.4728609, (0 missing)

Node number 2: 56 observations
 mean=0.8570631, MSE=0.2400441

Node number 3: 15 observations
 mean=2.162431, MSE=0.6020951
```

nCPA

```
         cp       r2 mean left mean right pperm       5%       25%       50%       75%       95%
```

[1,] 0.6420038 0.4728609 2.162431 0.8570631 0.001 0.5880071 0.6420038 0.6444849 0.8865301 0.9014286

Section 3: Estuaries

E.M. Grantz, L.B. Massey, M.E. Evans-White, B.E. Haggard and J.T. Scott

EXECUTIVE SUMMARY

The Clean Water Act directs states to adopt water-quality standards for all water bodies, including estuaries. The push in water-quality standards over the last decade or more has been for the development of numeric nutrient criteria applicable to all water bodies, and protective of designated beneficial uses – primarily aquatic life and recreation. The EPA has published a series of guidance documents detailing technical approaches that can be used to develop numeric nutrient criteria, including frequency distribution of nutrient concentration data and stressor-response studies. There is a growing wealth of information on streams, rivers, lakes and reservoirs, although the literature available on nutrient criteria for estuaries is much more limited because this water body type is more complex, both in hydrology and nutrient dynamics. The EPA has released guidance for the development of estuarine nutrient criteria for stressor and response, based on classification by physical characteristics (i.e., possible grouping of estuaries, analyses of historical data, reference site approach, stressor-response relationships (i.e., modeling approaches). However, the EPA has not provided numerical guidance (based on frequency distribution of existing nutrient and biological conditions) for estuaries like that provided for lakes and reservoirs or streams and rivers. Nonetheless, states are directed to develop numeric nutrient criteria for estuaries and coastal water protective of beneficial uses, as defined by each state's regulatory agency. The development of numeric nutrient criteria will allow the states to manage nutrient enrichment of coastal waters and the associated effects on recreation and biological conditions.

The objective of Section 3 was to provide statistical support to the Texas Commission on Environmental Quality (TCEQ) to aid the development of numeric nutrient criteria for Texas estuaries by TCEQ. The first step in this process was to compile the geospatial and water quality data from 860 stations within 38 estuaries and 4 basins. These data were provided by TCEQ and collected under non-biased conditions. Following data reorganization and reduction, median values for each parameter were estimated at each station with 10 observations or greater and compiled into a median database. The parameters of primary concern were total phosphorus (TP), ortho-phosphate (PO_4-P; SRP), total nitrogen (TN), nitrate plus nitrite N (NO_x-N), and sestonic chlorophyll-a (chl-a). Frequency distributions including the minimum, 10th, 25th, 50th, 75th, and 90th percentiles, and maximum of these parameters were calculated for the general population at multiple spatial scales, specifically, by basin, eco-region levels III and IV, and basin by level III eco-region. Frequency distributions are presented in Section 3.1 and were intended to provide TCEQ with the percentile estimates for estuaries that were estimated as tools for setting nutrient criteria by the USEPA for other water body types.

States are progressing to development of nutrient criteria, but questions remain regarding the legitimacy of promulgating one numeric criterion across areas that may contain multiple basins, various eco-regions, and a myriad of land uses. Section 3.2 provides analyses of potential geospatial variability in median total nutrient (TP and TN), chl-a, and Secchi transparency for Texas estuaries using classification and regression tree (CART) and non-parametric changepoint analysis (nCPA). Geospatial variables included land use/land cover (LULC) categories, permitted municipal wastewater treatment (WWTP) plant discharge, and salinity. For Texas estuaries, geospatial models based on salinity provided the most effective grouping schemes for TP concentration, which decreased with increasing salinity, and Secchi transparency, which increased with increasing salinity. Models for TN concentration, which decreased with increasing salinity, were weak comparatively, but removing obvious high TN, high salinity outlier stations improved model strength by >3x. Salinity thresholds identified in CART ranged from 17-21 ppt. Among LULC categories, CART analyses indicated a threshold in TN concentration at 5% wetland land cover and a threshold in chl-a fluoro at 61% developed+agricultural land. Statistically significant models using permitted municipal WWTP discharge were only found for TP. These models had low explanatory power relative to other possible geospatial models and identified WWTP discharge effects on TP concentration at the highest loading rates only.

The frequency distribution approach should be used in conjunction with other statistically based methods that evaluate stressor-response relationships in aquatic systems. Section 3.3 provides analyses of potential nutrient thresholds (median TP, TN) to biological response (median chl-a measured spectrophotometrically and fluorometrically, and Secchi transparency) in Texas estuaries. The total nutrient thresholds identified for biological response variables ranged from 0.068 to 0.25 mg/L TP and 0.86-1.3 mg/L TN. The lowest and highest TN and TP thresholds consistently corresponded with analyses using Secchi transparency and chl-a spec, respectively, as response variables. While TN thresholds identified for all response variables consistently fell within a narrow range, TP thresholds varied by approximately 4x in Texas estuaries depending upon the response variable. The TP threshold identified for chl-a spec was in fact approximately 2x greater than any other TP threshold identified for any water body type in this study. Analyses using chl-a spec as a response variable, particularly chl-a spec vs. TP were likely affected by a high percentage of censored observations in the chl-a spec data. Obtaining reliable threshold estimates for chl-a spec in Texas estuaries may require incorporating censored data analysis techniques for estimating measures of central tendency.

INTRODUCTION

The Clean Water Act (CWA) directs states to adopt water-quality standards for all water bodies, including estuaries. The push in water-quality standards over the last decade or more has been for the development of numeric nutrient criteria applicable to all water bodies, and protective of designated beneficial uses – primarily aquatic life and recreation. The EPA has published a series of guidance documents detailing technical approaches that can be used to develop numeric nutrient criteria, including frequency distribution of nutrient concentration data (EPA, 2000) and the preferred stressor-response studies (EPA, 2010). There is a growing wealth of information on streams, rivers, lakes and reservoirs, although the literature available on estuaries and nutrient criteria is much more limited because of this water body type is more complex in nature, both in hydrology and nutrient dynamics.

The hydrology of estuaries is complex resulting from the mixing and density stratification of fresh and marine waters, and the variation in mixing play a role in nutrient dynamics and biological response. The sources of nutrients to estuaries are the surrounding landscape, tidal streams, large inland rivers draining to marine systems, ground water and even atmospheric deposition; these external sources may lead to accumulation within estuarine systems providing for internal cycling of nutrients and oxygen demand. However, nutrient inputs to estuaries are closely link to the freshwater inputs, including tidal streams and larger inland rivers draining into these systems. These interactions and myriad of processes make establishing a link between stressors (i.e., nutrients) and response variables (e.g., oxygen concentration, sestonic or benthic algae, water clarity, etc.) difficult at times.

The EPA has released guidance for the development of estuarine nutrient criteria for stressor and response (EPA, 2001), based on these selected elements:

- classification by physical characteristics (i.e., possible grouping of estuaries)
- analyses of historical data
- reference site approach
- stressor-response relationships (i.e., modeling approaches)

The guidance also focuses on examining the information and proposed nutrient criteria by a panel of regional, federal, state and tribal experts and on determining the consequences of the criteria both upstream and downstream. However, the EPA has not provided numerical guidance [based on frequency distribution of existing nutrient and biological conditions] for estuaries like that provided for lakes and reservoirs (EPA 2000b) or streams and rivers (EPA 2000a). Nonetheless, states are directed to develop numeric nutrient criteria for estuaries and coastal water protective of beneficial uses, as defined by each state's regulatory agency. The development of numeric nutrient criteria will allow the states to manage nutrient enrichment of coastal waters and the associated effects on recreation and biological conditions.

The effects of nutrient enrichment share some similarities with freshwaters, but the considerations generally focus on the development of harmful algal blooms (i.e., algal toxins), hypoxic conditions in coastal waters and the impact on marine biology (i.e., accelerated eutrophication and oxygen demand), loss of submerged and shore-line vegetation native to coastal waters. Estuaries might also not show an increase in primary productivity (carbon fixation) with increased nutrient enrichment, because the productivity shifts from benthic algae to sestonic organisms (McGlathery et al. 2007). However, nutrient loads are positively correlated with the anoxic volume in coastal waters (Kemp et al. 2005, Conley et al. 2009, Brietberg et al. 2009), showing the nutrient enrichment reduces oxygen concentrations in overlying water especially when density stratification occurs (Nestlerood 1998, Stow et al. 2005). The literature on estuaries has generally focused on the relationship between dissolved oxygen, nutrients and biological conditions, and few studies have employed threshold analysis or even fewer identified nutrient thresholds. The associated effects of nutrient enrichment also influence the social and economic values, such as recreational opportunities, cultural uses and marine fisheries.

The objectives of this chapter are:

1) to discuss the frequency distributions of median nutrient concentrations and response variable for Texas estuaries at various spatial scales from data acquired from the Texas Commission on Environmental Quality (TCEQ),
2) to explore ways to group estuaries based on chemical and physical conditions, such as salinity and watershed-coastal characteristics, for the development of nutrient criteria,
3) to identify nutrient threshold values associated with changes in the magnitude or variability of commonly measured biological response variables for Texas estuaries.

3.1: ESTUARIES DATABASE DEVELOPMENT, MEDIAN CALCULATION, AND FREQUENCY DISTRIBUTIONS

Methods

Water Quality Database

Data Acquisition, Compilation and Reduction. TCEQ provided a database of water quality data collected from 1968 to 2012 from estuaries along the Texas coastline. The collected data was from 860 stations from 34 estuaries was divided among three Microsoft Excel workbooks. The data described 116 estuary characteristics and water quality parameters including nutrients, sediments, transparency, physico-chemical parameters, as well as others.

For the purposes of advanced statistical analyses conducted during this project, only data collected under specific monitoring type codes (as decided by TCEQ) and from 2000 to 2010 was used. Therefore, the database was sorted and any data collected before calendar year 2000 or after 2010 was removed. Data collected under the monitoring type code Biased Flow (BF) was also removed since data collected under this circumstance were not necessarily representative of baseline water quality conditions. The data received from TCEQ were output to a single column format within the files, so the data were reorganized into a useable format. The data was sorted by Basin ID and a new Microsoft Excel worksheet was created for each individual basin. Each basin worksheet was then restructured using the pivot table function in Microsoft Excel so that each parameter and the associated data were unique to an individual column; a portion of this process was accomplished with a Microsoft Excel Macro (see Appendix 1.1 for Excel Macro code). Any estimated data points (i.e., those reported with a < or >) were flagged and used in the database without the associated qualifying sign. The data were flagged using a Microsoft Excel Macro (Appendix 1.2).

Several additional parameters were calculated from the original data provided. Nitrate plus nitrite and total N (TN) were calculated if the necessary N species were provided by TCEQ in the original data file. In addition, diel change (i.e., 24 hour maximum minus 24 hour minimum) was calculated for dissolved oxygen, temperature, conductivity, pH, and turbidity. The additional parameters were added to each station worksheet using a Microsoft Excel Macro (Appendix 1.3).

Due to the volume of data provided, several parameters were removed from the median database because of lack of data and duplication of parameters, or because TCEQ indicated that the parameter could be removed from the database.

Median and Frequency Distribution Calculations. For this study, frequency distribution and, subsequently, stressor-response analyses were conducted on station medians in order to focus on broadly applicable regional and statewide trends. Because each estuary in Texas was not equally represented in the raw water quality dataset, conducting statistical analyses on medians removes potential site-specific bias for sites that are over- or under-represented in the raw dataset. Furthermore, biological response and nutrient stressor data did not always overlap in the raw data. Conducting analyses with median values allowed comparison of long-term trends in biological and nutrient data for these stations. Median values of each parameter were calculated for each Station ID using a Microsoft Excel Macro (Appendix 1.4). Median values were calculated based on at least 10 data points, i.e. no medians were calculated if less than 10 data points were available for a given parameter at a given station. The calculated medians for each Station ID were then compiled into one database using a Microsoft Excel Macro (Appendix 1.5). This database was merged with the GIS and LULC data and used in advanced statistical analysis.

Frequency distributions (minimum value, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles and maximum value) for water quality parameters TP (TCEQ parameter code 00665), TN (calculated parameter code 00600C; TCEQ parameter code 00625 + 00630, 00625 + 00593 or 00625 + 00615 + 00620), NO_x-N (calculated parameter 00630C; TCEQ parameter code 00630, 00593 or 00615 + 00620), PO_4-P (TCEQ calculated parameter code 00671C; TCEQ parameter code 00671 or 70507), and chl-a (TCEQ parameter code 70953) were calculated using Microsoft Excel. For this study, a parameter combining chl-a measured spectrophotometrically (parameter code 32211; chl-a spec) and chl-a fluoro was not created due to inconsistencies between the methods (Laurie Eng, personal communication). Data were more complete and censorship was less of a concern for chl-a fluoro than for chl-a spec. Spectrophotometric chl-a data were commonly censored at a relatively high detection limit (10 µg/L). Analysis exploring the effects of censored data on chl-a spec median calculation in Texas reservoirs indicated that, when censored data exceeded 16% of the raw data for a station, chl-a medians (i.e. the 50th percentile) were increasingly overestimated when the detection limit was substituted for censored observations as the level of censoring increased because the median could not be calculated to be a value below the detection limit (see Section 2.3). This censored data effect would have been magnified further when considering very low percentiles in the frequency distribution, such as the 25th percentile. Therefore, frequency distributions for sestonic chl-a were only calculated for the fluorometric method in this study.

Geospatial Database

A geospatial database contained within a Microsoft Excel file was provided by TCEQ that identified land use and land cover data for the water quality stations located within estuaries included in this study. The geospatial descriptors were provided for the drainage basin, which was defined as having an upstream boundary constrained by the nearest upstream reservoir. The descriptors included percent open water, developed-open, developed-low intensity, developed- medium intensity, developed-high intensity, barren land, deciduous forest, evergreen forest, mixed forest, shrub/scrubland, grassland/herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands. These descriptors were reduced to five categories including percent developed (i.e., open, low intensity, medium intensity, barren land), forest (i.e., deciduous, evergreen, mixed and shrubland), agriculture (i.e., grassland/herbaceous, pasture/hay, cultivated crops), developed plus agriculture (i.e., open, low intensity, medium intensity, barren land, grassland/herbaceous, pasture/hay, cultivated crops) and wetlands (i.e., woody and emergent herbaceous wetlands). Additional geospatial information for each site was provided including drainage area, slope, municipal discharges, basin ID, level III ecoregion ID, and level IV ecoregion ID.

Data Quality Assurance and Control

Data quality checks were employed frequently throughout the database reorganization and data calculation processes. The original source files were maintained in an unaltered form, and subsequent changes to each database were saved under unique file names. Data transferred from one file to the next were checked for accuracy by comparing first and last rows and the row count between files. In addition, when calculations were preformed, including manual calculations and those calculated using Microsoft Excel Macros, at least 10 percent of calculations were checked for accuracy following the secondary data quality assurance project plan (QAPP).

Results and Discussion

The Texas Coast portion of the Gulf of Mexico includes nine major bay systems, and estuaries span four basins; however, data was very limited from Basins 7, 11, and 13. Therefore, 25th percentiles could only be calculated for Basin 24; 50th percentiles (medians) are provided for Basins 7 and 11 (Table 3.1.1). Median total phosphorus concentrations were calculated from 204 stations located in Basin 24, and the 25th percentile of the median TP concentration was 0.08 mg/L. Median PO_4-P concentrations were calculated from 200 stations in Basin 24, and the 25th percentile of the median PO_4-P concentration was 0.06 mg/L. Station 13589, Bayport Channel at Turning Basin, had the greatest median concentration for both TP and PO_4-P.

Less TN data was available compared to phosphorus, and 158 estuary stations located in Basin 24 contributed to the frequency distribution of median TN concentrations. The 25th percentile of the median TN concentration was 0.76 mg/L, and the greatest median concentration was 2.02 mg/L which was observed at Station 13450, Baffin Bay at CM 14. Median NO_X-N concentrations were calculated based on data from 203 estuary stations. The 25th percentile of these stations in Basin 24 was 0.04 mg/L. The greatest media concentration observed was 0.94 mg/L which was observed at Station 17923, Upper San Jacinto Bay near Hiwires.

Limited chl-a data was available from the estuary stations. Frequency distributions were calculated from the 70 stations in Basin 24 with adequate data. The 25th percentile of the median chl-a concentrations was 4.96 µg/L, and the maximum median observed was 36.9 µg/L at Station 13335, Clear Lake at CM 17.

Table 3.1.1. Frequency distribution of median nutrient and chlorophyll-a concentrations from estuaries among basins in Texas, 2000-2012; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
7	1	--	--	--	0.080	--	--	--
11	1	--	--	--	0.060	--	--	--
13	0	--	--	--	--	--	--	--
24	204	0.050	0.060	0.080	0.120	0.179	0.254	0.755

Total Nitrogen (TN; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
7	1	--	--	--	0.73	--	--	--
11	1	--	--	--	0.78	--	--	--
13	0	--	--	--	--	--	--	--
24	158	0.55	0.65	0.76	0.92	1.02	1.25	2.02

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
7	1	--	--	--	0.11	--	--	--
11	1	--	--	--	0.17	--	--	--

13	0	--	--	--	--	--	--	--
24	203	0.02	0.04	0.04	0.09	0.17	0.31	0.94

Ortho-Phosphate (PO_4-P; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
7	1	--	--	--	0.060	--	--	--
11	1	--	--	--	0.110	--	--	--
13	0	--	--	--	--	--	--	--
24	200	0.015	0.040	0.060	0.095	0.180	0.230	0.810

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
7	1	--	--	--	5.57	--	--	--
11	1	--	--	--	12.8	--	--	--
13	0	--	--	--	--	--	--	--
24	70	3.00	4.00	4.96	6.03	9.48	16.8	36.9

The entire coastal region is part of the Western Gulf Coastal Plain Level IIII ecoregion, and all estuary monitoring sites lie within this Level III ecoregion (Table 3.1.2). The frequency distributions calculated at the Level III ecoregion include the stations within Basin 24 as well as Station 10683 in Basin 7 and Station 11498 in Basin 11. The addition of these two stations did have much effect on the distribution at the Level III ecoregion; the25th percentile of the median TN concentration decreased from 0.76 mg/L at the Basin level to 0.75 mg/L while the 25th percentile of the median chl-a concentrations increased from 4.96 µg/L at the Basin level to 5.00 µg/L at the level III ecoregion.

Table 3.1.2. Frequency distribution of median nutrient and chlorophyll-a concentrations from estuaries among level III ecoregions in Texas, 2000-2012; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	206	0.050	0.060	0.080	0.120	0.176	0.253	0.755

Total Nitrogen (TN; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	160	0.55	0.65	0.75	0.92	1.01	1.25	2.02

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	205	0.02	0.04	0.04	0.09	0.17	0.31	0.94

Phosphate (PO_4-P; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	202	0.015	0.040	0.060	0.095	0.180	0.230	0.810

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	72	3.00	4.04	5.00	6.03	9.74	16.5	36.9

3.2 GROUPING ESTUARY STATIONS WITH SIMILAR NUTRIENT AND BIOLOGICAL CONDITIONS BY THRESHOLDS IN GEOSPATIAL VARIABLES

Methods

We conducted Classification and Regression Tree (CART) analyses on the median database for Texas estuaries (described in Section3.1) to group stations exhibiting similar nutrient conditions by watershed attributes. The focus (dependent) nutrient variables of these analyses were median TP and TN concentrations (TCEQ parameter codes 00665 and 00600C, respectively). The focus biological variables were chlorophyll- measured both spectrophotometrically and fluorometrically and Secchi transparency (TCEQ parameter codes 32211, 70953, and 00078C, respectively). The watershed attributes considered as predictor (independent) variables were divided into three categories: 1) watershed land use/land cover, 2) permitted municipal waste water treatment plant flow, and 3) salinity. Land use/land cover was further divided into percent developed, percent agriculture, percent developed + agriculture, percent forested, and percent wetland. Permitted municipal waste water treatment plant discharge was used as a predictor variable both unweighted (mgd) and weighted by watershed area (mgd/km^2). Texas estuaries are located almost exclusively within a single level III EPA ecoregion (34-Gulf Coastal Plains) and a single basin (24-Bays and Estuaries), as defined by TCEQ. Therefore, geospatial analysis based on regional groupings was not possible. However, salinity was used as a proxy for regional geospatial variability that was assumed to be related to flow regime (freshwater inflow, WWTP discharge, and precipitation) in the contributing basin. The hypothesis was that TP, TN and chlorophyll-a concentrations would exhibit an inverse relationship with salinity, while Secchi transparency would be directly related to salinity and increase with increasing salinity.

CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (The models with the greatest explanatory power were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance (p_{perm}<0.05) and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value (see Section 3.1).

A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.10. In Appendix 3.2, the statistical code and raw output generated for each CART and nCPA analysis conducted for this study on geospatial variability in Texas estuaries has been compiled.

Results and Discussion

Land use/Land cover

Land use/land cover categories were weak predictors for TP and TN concentrations. The %Developed LULC was the strongest predictor among LULC categories for TP (Fig. 3.2.1A), and CART indicated that TP concentrations were higher on average above a threshold of 38% developed land. For TN, %Wetland LULC was the strongest predictor among LULC categories (Fig. 3.2.1B). On average, TN concentrations were higher below a threshold of 5.0% wetland cover. Wetlands are known hotspots in the landscape for physical and biogeochemical removal processes (Mitch and Gosselink 2007), such as denitrification, the transformation of reactive NO_x-N to inert N_2 gas, which removes N from the system via diffusion into the atmosphere. Therefore, this finding is congruent with established ecological theory, and the identified threshold in wetland cover was in range with recommended wetland to watershed surface area ratios recommended for constructed wetlands. For both biological variables, CART identified the combined percent developed and agricultural cover as a predictor variable. For chl-a fluoro, approximately 60% developed+agricultural land was a threshold that resulted in biological change (doubling of chlorophyll-a concentrations of average; Fig. 3.2.1C). For Secchi transparency, the threshold in developed+agricultural land was lower, or approximately 40% (Fig. 3.2.1D). No statistically significant splits for spectrophotometric chlorophyll-a were identified for Texas estuaries.

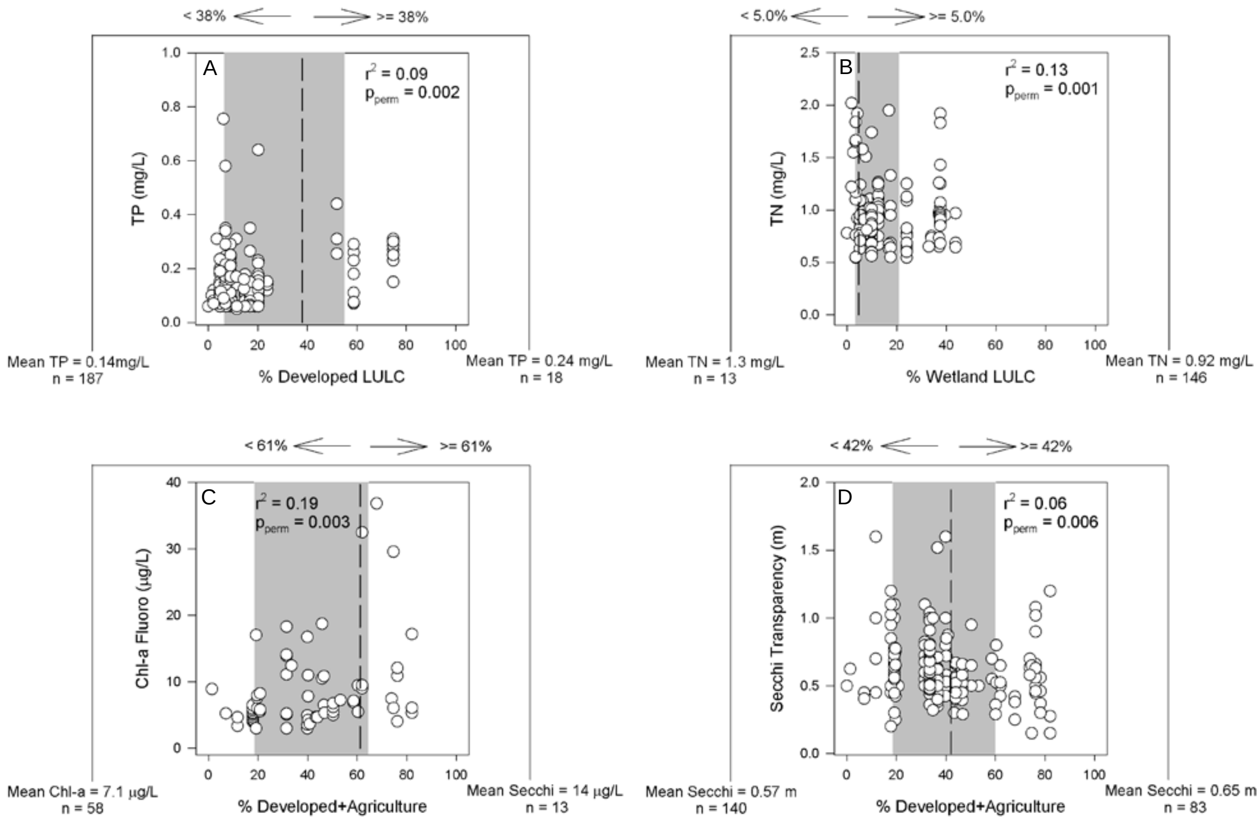


Figure 3.2.1. The relationship between (A) median total phosphorus (TP) and percent developed land use, (B) median total nitrogen (TN) and percent wetland land use, and (C) chl-a fluoro and (D) Secchi transparency and percent developed+agriculture across Texas estuaries showing thresholds based on classification and regression tree analysis (CART).

Permitted municipal WWTP Flow

For TP, geospatial models based on permitted municipal WWTP flow performed similarly both unweighted and weighted for watershed area (Fig. 3.2.2A-B). Model strength was slightly higher for TP vs. weighted discharge (r^2 = 0.05 vs. 0.04), but both models had very low explanatory power. Furthermore, CART identified a high threshold permitted municipal WWTP discharge rate, and this changepoint split only a relatively small number of observations from the rest of the data. This finding indicates that the few potential WWTP effects that were observable in the dataset only occurred at the highest loading rates. No statistically significant splits were identified for TN and chlorophyll-a concentrations or Secchi transparency using either unweighted or weighted permitted municipal WWTP flow.

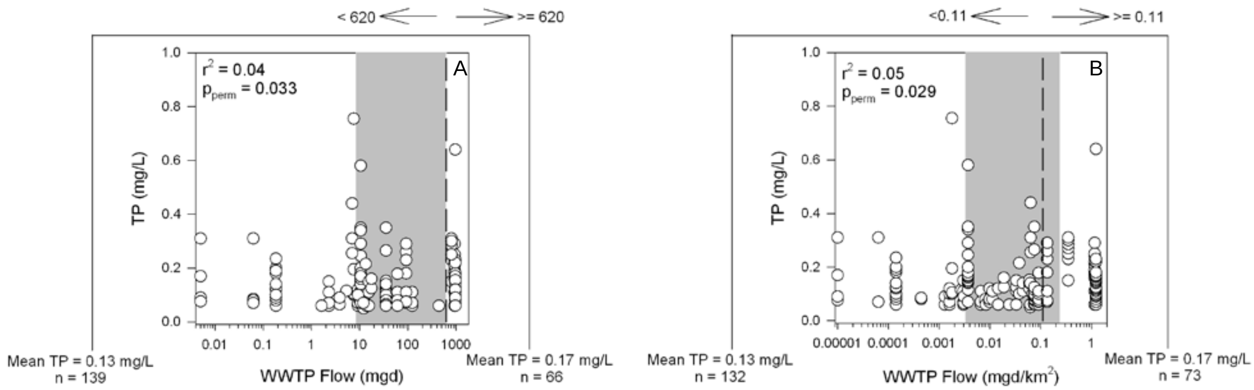


Figure 3.2.2. The relationship between median total phosphorus (TP) and waste water treatment plant (WWTP) flow before (A) and after (B) weighting by watershed area across Texas estuaries showing thresholds based on classification and regression tree analysis (CART).

Salinity

Median TP and salinity were strongly related for estuaries (Fig. 3.2.3A), indicating that, on average, TP concentrations were higher at low salinity (<17 ppt). The CART model for median TN vs. salinity for estuaries was weaker (Fig. 3.2.3B), but indicated a similar salinity threshold (17 ppt) and relationship between nutrient concentrations and salinity. Six extreme "outlier" stations with TN > 1.5 mg/L and salinity > 30 ppt were evident in CART analysis of TN vs. salinity for estuaries. The estuary ID's for these stations were identified to determine if these outliers had obvious geospatial similarities (Table 3.2.1). Each outlier station was located within Laguna Madre or South Texas estuaries connected to Laguna Madre. The basins exporting flow to these estuaries are arid, resulting in more saline conditions relative to estuaries where contributing basins are located within climates with greater precipitation. Removing the six extreme outlier sites from the model, improved model strength and significance (Fig. 3.2.3C). Threshold salinity increased to 21 ppt, but was still in range with the salinity threshold identified for TP. These findings suggest that the basin contributing flow to the estuary and associated flow regime may also be an important geospatial predictor for nutrient concentrations in Texas estuaries, particularly for TN. For South Texas estuaries, estuary nutrient concentrations may be less related to watershed exports than in other Texas estuaries. Therefore, for these systems, salinity may not be as reliable a predictor for nutrient concentrations. Though all six outlier sites were within South Texas estuaries, most South Texas estuary sites were not outliers in the TN vs. salinity model. These findings may be most applicable to sites that often exhibit elevated salinity. Future work on geospatial grouping of Texas estuaries should explore potential geospatial differences between the low and high salinity groups identified in CART analyses and refine the methods employed here for grouping estuary stations based on characteristics of the contributing basin.

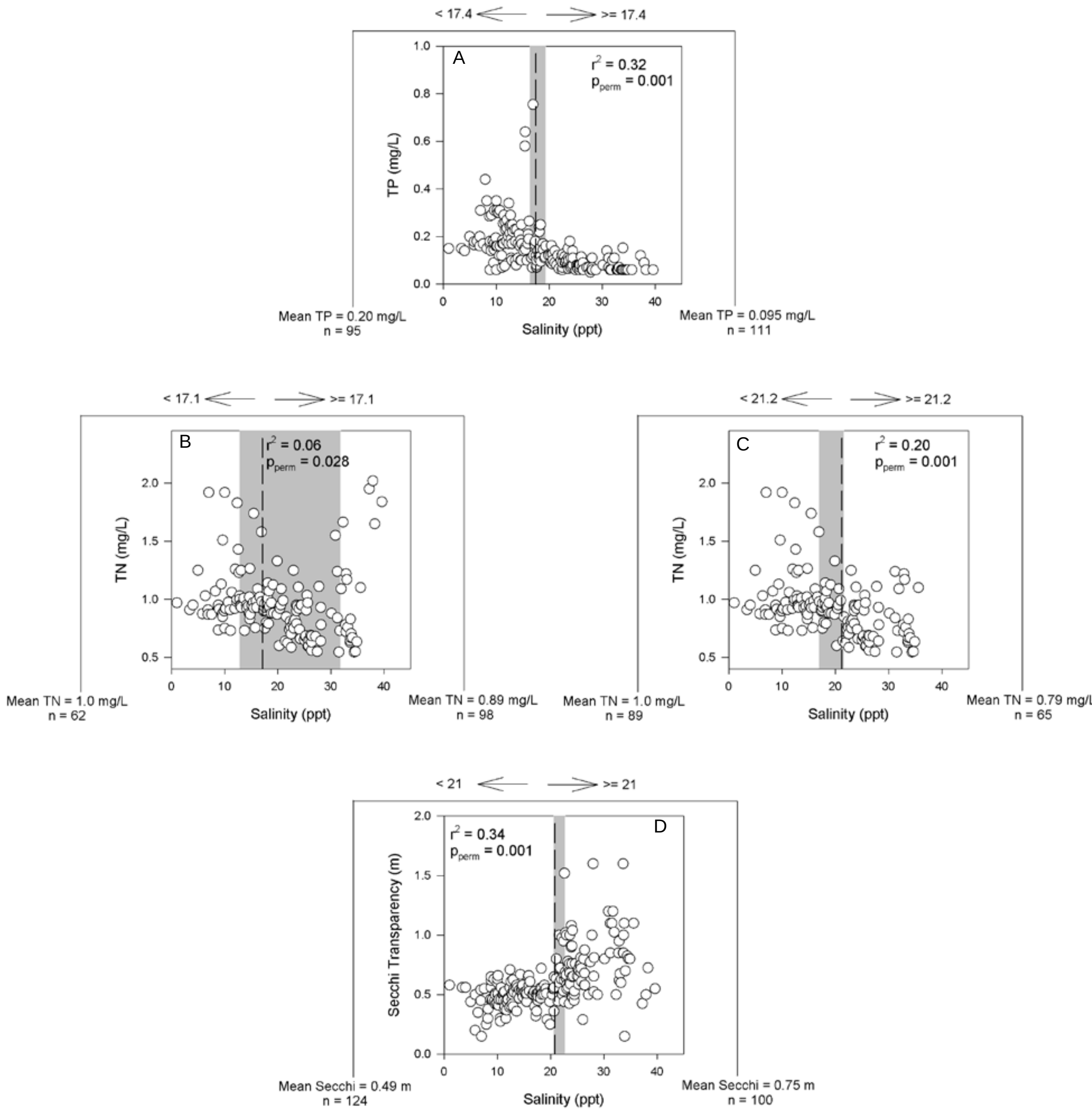


Figure 3.2.3. The relationship between (A) median total phosphorus (TP) and (B) median total nitrogen (TN) and salinity across Texas estuaries showing thresholds based on classification and regression tree analysis (CART). CART analyses of the relationship between TN and salinity was also conducted after removing six extreme outlier sites (C). Finally, CART analysis of the relationship between Secchi transparency and salinity was also carried out (D).

Table 3.2.1. List of potential outlier stations in classification and regression tree (CART) analysis of the relationship between total nitrogen (TN) and salinity across estuaries in Texas.

Station	Estuary ID	Description
13440	Oso Bay	OSO BAY AT PADRE ISLAND DR
13445	Laguna Madre	LAGUNA MADRE ICWW/BIRD ISLAND
13446	Laguna Madre	LAGUNA MADRE GIWW CM 129
13448	Laguna Madre	LAGUNA MADRE AT GIWW
13452	Baffin Bay/Alazan Bay/Cayo Del Grullo/LagunaSalada	BAFFIN BAY AT CM 36
13459	South Bay/Brownsville Ship Channel	SOUTH BAY NEAR CLARK ISLAND

Finally, Secchi transparency was also strongly related to salinity in Texas estuaries (r^2=0.34), and exhibited a direct relationship with salinity (Fig. 3.2.4), increasing as salinity increased. Above a threshold salinity of 21 ppt, Secchi transparency was ~50% greater on average. This threshold was in range with the salinity thresholds identified for TP and TN concentrations. No statistically significant salinity thresholds were identified for chlorophyll-a concentrations.

Summary of Geospatial Analysis for Texas Estuaries

A summary of the strongest CART models for median TP, TN, and chlorophyll-a concentrations, and Secchi transparency, is available in Table 3.2.2. For Texas estuaries, models for grouping stations geospatially with the greatest strength were based on salinity gradients, especially for TP and Secchi transparency. After identifying and removing extreme outlier sites, the model for TN concentration vs. salinity also exhibited good predictive power. Models for TP and TN concentration and Secchi transparency vs. salinity identified a salinity threshold for nutrients in the range of 17-21 ppt. In general, geospatial grouping schemes based on LULC or WWTP flow were not effective for Texas estuaries, though the percent developed+agricultural land was a good predictor for chl-a fluoro.

Table 3.2.2. Summary of CART models for estuaries with the greatest statistical power in each geospatial category (LULC, WWTP Flow, and Region). NS= Not significant.

Parameter	Geospatial Category	Predictor	Threshold	Model r^2
Total Phosphorus	LULC	% Developed	38%	0.06
	WWTP Flow	Weighted by watershed area	0.11 mgd/km^2	0.05
	Salinity	Salinity	17 ppt	0.32
Total Nitrogen	LULC	% Wetland	5.0%	0.13
	WWTP Flow	NS	NS	NS
	Salinity	Salinity (outliers removed)	21 ppt	0.20
Chl-a Spectrophotometric	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Salinity	NS	NS	NS
Chl-a Fluorometric	LULC	%Developed+Agriculture	61%	0.19
	WWTP Flow	NS	NS	NS
	Salinity	NS	NS	NS
Secchi Transparency	LULC	%Developed+Agriculture	42%	0.06
	WWTP Flow	NS	NS	NS
	Salinity	Salinity	21 ppt	0.34

3.3: STRESSOR-RESPONSE ANALSIS ON THE ESTUARY MEDIAN WATER QUALITY DATABASE

Methods

We conducted CART analyses on the median database for estuaries (described in Section 3.1) to identify thresholds in nutrient concentrations that resulted in measurable changes in common biological responses. The biological (dependent) variables included in the analyses were: median Secchi depth (m), median 24 hour dissolved oxygen (DO) flux, median chlorophyll-a (chl-a) measured with spectrophotometry, and median chl-a measured with fluorometry. The nutrient (independent) variables included in the analysis were median total phosphorus (TP; 00665) and median total nitrogen (TN; 00600C).

CART analysis is a form of data reduction that aims to: 1) quantify thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables, and 2) identify hierarchical structure in independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian et al. 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). The models with the greatest explanatory power were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance ($p_{perm}<0.05$) and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value. Because CART analysis involves recursive partitioning, models may sometimes be over-fit (i.e. too many independent variables that decrease the statistical rigor of final model). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model crossvalidations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.10. In Appendix 3.3, the statistical code and raw output generated for each CART and nCPA analysis conducted on stressor-response relationships in Texas estuaries has been compiled.

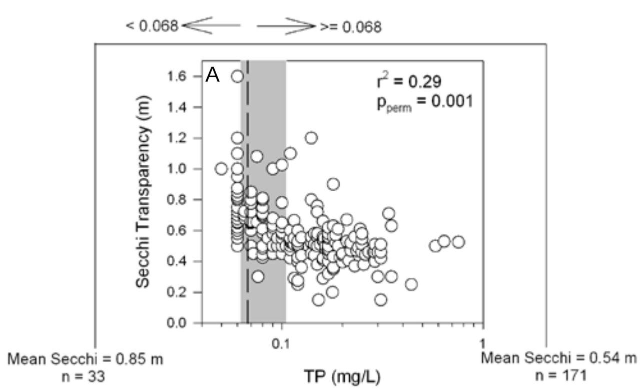
Results and Discussion

Secchi transparency

For Secchi transparency, thresholds in TP and TN concentrations were found for Texas estuaries (Figs. 3.3.1A-B., respectively. On average, Secchi transparency was more than 50% higher when TP concentration was below 0.068 mg/L and was approximately 25% higher when TN concentrations were below 0.85 mg/L. Model strength for Secchi transparency vs. TP was 2x greater than for Secchi transparency vs. TN.

Chlorophyll-a

For chlorophyll-a measured both spectrophotometrically (Figs. 3.3.2A-B) and fluorometrically (Figs. 3.3.2C-D), thresholds in TP and TN concentrations were found for Texas estuaries. On average, spectrophotometric chlorophyll-a concentrations were 30-40% greater when TP concentrations exceeded 0.25 mg/L and TN concentrations exceeded 1.3 mg/L. The fact that average chlorophyll-a concentration for stations below the nutrient changepoints was equal to 10 µg/L, a common detection limit, may indicate that the spectrophotometric chlorophyll-a vs. nutrient analyses may have been strongly influenced by censored data. Increases in average fluorometric chlorophyll-a above nutrient thresholds exceeded 50% for both TN and TP. Nutrient thresholds identified for fluorometric chlorophyll-a were also lower than those identified for spectrophotometric chlorophyll-a, especially for TP (0.11 mg/L vs. 0.25 mg/L). Model strength for chlorophyll-a vs. TN consistently exceeded model strength for chlorophyll-a vs. TP.



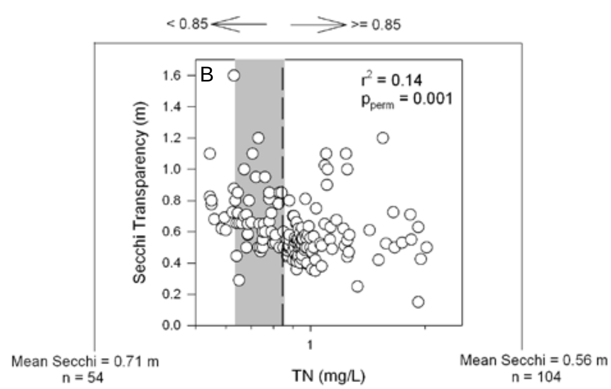


Figure 3.3.1. The relationship between Secchi transparency and (A) total phosphorus (TP) and (B) total nitrogen (TN) across Texas estuaries showing thresholds based on classification and regression tree analysis (CART).

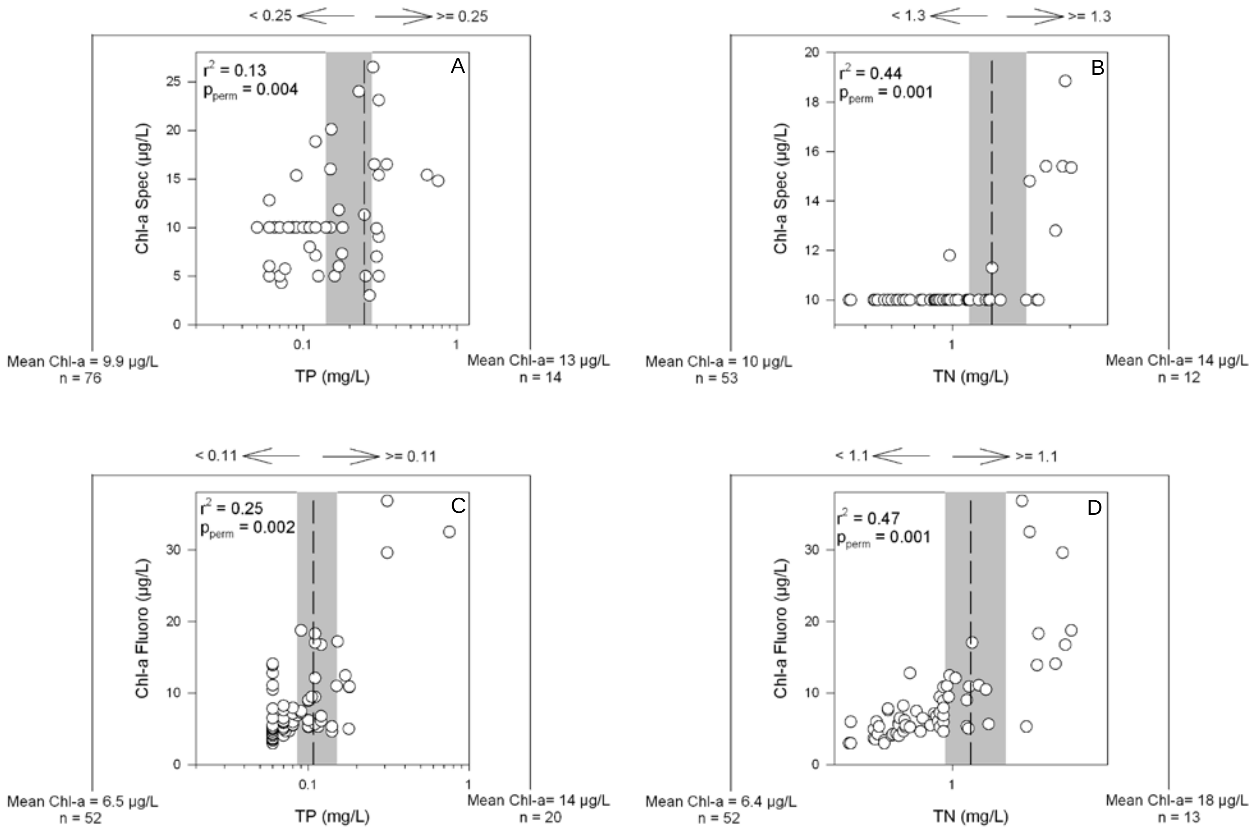


Figure 3.3.2. The relationship between chlorophyll-a measured spectrophotometrically (Chl-a Spec) and (A) total phosphorus (TP) and (B) total nitrogen (TN), as well as between chlorophyll-a measured fluorometrically (chl-a fluoro) and (C) total phosphorus (TP) and (D) total nitrogen (TN) across Texas estuaries showing thresholds based on classification and regression tree analysis (CART).

Summary of Stressor-Response Analysis for Texas Estuaries

Total nutrients were consistently strong predictors for biological response in Texas estuaries. For Secchi transparency, the CART model based on TP concentration had greater predictive power than that based on TN (r^2=0.29 vs. r^2=0.14). This trend was reversed for chl-a measured using both methods. The nutrient thresholds identified in these analyses ranged from 0.068-0.25 mg/L TP and 0.85-1.3 mg/L TN. The lowest and highest nutrient thresholds were consistently found for Secchi transparency and chl-a spec, respectively. While TN thresholds identified for all response variables fell within a narrow range, TP thresholds varied by approximately 4x in Texas estuaries depending upon the response variable. The TP threshold identified for chl-a spec was in fact approximately 2x greater than any other TP threshold identified for any water body type in this study. Analyses using chl-a spec as a response variable, particularly chl-a spec vs. TP were likely affected by a high percentage of censored observations in the chl-a spec data (see Fig.3.3.3A). Obtaining reliable threshold estimates for chl-a spec in Texas estuaries may require incorporating censored data analysis techniques for estimating metrics of central tendency.

REFERENCES

Breitburg, D.L., D.W. Hondorp, L.W. Davias, and R.J. Diaz. 2009. Hypoxia, nitrogen, and fisheries: Integrating effects across local and global landscapes. Annual Reviews in Marine Science 1:329-350.

Conley, D.J., G.E. Likens, D. Buso, L. Saccone, S.W. Bailey, C.E. Johnson. 2008. Deforestation causes increased dissolved silicate losses in the Hubbard Brook Experimental Forest. Global Change Biol. 14: 2548-2554.

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

East T L, Sharfstein B. 2006. Development of a decision tree model for the prediction of the limitation potential of phytoplankton in Lake Okeechobee, Florida, USA. Arch. Hydrobiol. 2006;165:127-144.Environmental Protection Agency (EPA). 2000. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002.

Environmental Protection Agency (EPA). 2000a. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002.

Environmental Protection Agency (EPA). 2000b. Nutrient criteria technical guidance manual: Lakes and reservoirs. EPA-822-B-00-001.

Environmental Protection Agency (EPA). 2001. Nutrient criteria technical guidance manual: estuarine and coastal marine waters. EPA-822-B-01-003.

Environmental Protection Agency (EPA). 2010. Using stressor-response relationships to derive numeric nutrient criteria. EPA 820-S-10-001.

Kemp, W.M., W.R. Boynton, J.E. Adolf, D.F. Boesch, W.C. Biocourt, G. Brush, J.C. Cornwell, R.R. Fisher, P.M. Gilbert, J.D. Hagy, L.W. Harding, E.D. Houde, D.G. Kimmel, W.D. Miller, R.I.E. Newell, M.R. Roman, E.M. Smith, J.C. Stevenson. Eutrophication of Chesapeake Bay: historical trends and ecological interactions. 2005. Marine Ecol Prog Ser. 303:1-29.

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(3):998-1008.

King RS, and CJ Richardson. 2003. Integrating bioasessment and ecological risk assessment: An approach to developing numerical water-quality critera. Environmental Management. 31:795-809. doi:10.1007/s00267-002-0036-4Longing SD, Haggard BE. 2010. Distributions of median nutrient and chlorophyll concentrations across the Red River Basin, USA. Journal of Environmental Quality 39:1966-1974.

McGlathery K.J., Sundback, K, I.C. Anderson. 2007. Eutrophication in shallow coastal bays and lagoons: the role ofplants in the coastal filter. Marine Ecology Progress Series. 348: 1-18.

Mitch, W.J. and J.G. Gosselink. 2007. Wetlands. John Wiley and Sons, Hoboken, NJ.

Qian SS, King RS, Richardson CJ. 2003. Two methods for the detection of environmental thresholds. Ecological Modeling 166:87-97.Robertson et al. 2006

Stow CA, Qian SS, and Craig JK. 2005. Declining threshold for hypoxia in the Gulf of Mexico. Environ Sci Technol 39: 716–23.

Urban, D.L. (2002) "Classification and regression trees" In: McCune, B., Grace, J.B. eds. , Analysis of Ecological Communities, MjM Software Design, Gleneden Beach, Oregon, USA, pp 221-231.

APPENDICES

Appendix 3.1. Frequency distributions of estuary data for basin x level III and level IV Ecoregions.

Table A3.1.1. Frequency distribution of median nutrient and chlorophyll-a concentrations from estuaries among basin by level III ecoregions in Texas, 2000-2012; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
7-34-Western Gulf Coastal Plain	1	--	--	--	0.080	--	--	--
11-34-Western Gulf Coastal Plain	1	--	--	--	0.060	--	--	--
24-34-Western Gulf Coastal Plain	204	0.050	0.060	0.080	0.120	0.179	0.254	0.755

Total Nitrogen (TN; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
7-34-Western Gulf Coastal Plain	1	--	--	--	0.73	--	--	--
11-34-Western Gulf Coastal Plain	1	--	--	--	0.78	--	--	--
24-34-Western Gulf Coastal Plain	158	0.55	0.65	0.76	0.92	1.02	1.25	2.02

Nitrite Plus Nitrate-Nitrogen (NO_X-N; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
7-34-Western Gulf Coastal Plain	1	--	--	--	0.11	--	--	--
11-34-Western Gulf Coastal Plain	1	--	--	--	0.17	--	--	--
24-34-Western Gulf Coastal Plain	203	0.02	0.04	0.04	0.09	0.17	0.31	0.94

Ortho-Phosphate (PO_4-P; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
7-34-Western Gulf Coastal Plain	1	--	--	--	0.060	--	--	--
11-34-Western Gulf Coastal Plain	1	--	--	--	0.110	--	--	--
24-34-Western Gulf Coastal Plain	200	0.015	0.040	0.060	0.095	0.180	0.230	0.810

Fluorometric Chlorophyll-a (Chl-a; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
7-34-Western Gulf Coastal Plain	1	--	--	--	5.57	--	--	--
11-34-Western Gulf Coastal Plain	1	--	--	--	12.8	--	--	--
24-34-Western Gulf Coastal Plain	70	3.00	4.00	4.96	6.03	9.48	16.8	36.9

Table A3.1.2. Frequency distribution of median nutrient and chlorophyll-a concentrations from estuaries among level IV ecoregions in Texas, 2000-2012; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	61	0.115	0.150	0.170	0.220	0.290	0.340	0.755
34b-S Subhumid Glf Cstl Prair	4	0.060	--	0.060	0.060	0.080	--	0.140
34d-Coastal Sand Plain	0	--	--	--	--	--	--	--
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	0.060	--	--	--
34g-Texas-Louisiana Cstl Marsh	20	0.060	0.069	0.088	0.155	0.180	0.181	0.200

34h-Mid-Coast Barr Isl C Marsh	104	0.050	0.060	0.072	0.100	0.120	0.159	0.230
34i-Laguna Madre Br Isl C Mrs	16	0.060	0.060	0.060	0.060	0.075	0.115	0.152

Total Nitrogen (TN; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	42	0.78	0.87	0.92	1.00	1.25	1.57	1.92
34b-S Subhumid Glf Cstl Prair	4	0.72	--	0.73	0.78	0.86	--	0.95
34d-Coastal Sand Plain	0	--	--	--	--	--	--	--
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	0.63	--	--	--
34g-Texas-Louisiana Cstl Marsh	17	0.73	0.73	0.75	0.92	0.98	1.02	1.03
34h-Mid-Coast Barr Isl C Marsh	81	0.55	0.63	0.69	0.88	0.95	1.04	1.55
34i-Laguna Madre Br Isl C Mrs	15	0.55	0.58	0.66	1.10	1.66	1.91	2.02

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	62	0.03	0.05	0.09	0.15	0.31	0.64	0.94
34b-S Subhumid Glf Cstl Prair	4	0.04	--	0.04	0.06	0.09	--	0.10
34d-Coastal Sand Plain	0	--	--	--	--	--	--	--
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	0.04	--	--	--
34g-Texas-Louisiana Cstl Marsh	18	0.04	0.04	0.06	0.10	0.16	0.24	0.25
34h-Mid-Coast Barr Isl C Marsh	104	0.02	0.04	0.04	0.05	0.15	0.25	0.55
34i-Laguna Madre Br Isl C Mrs	16	0.02	0.04	0.04	0.04	0.10	0.17	0.30

Phosphate (PO_4-P; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	60	0.050	0.090	0.120	0.160	0.203	0.241	0.810
34b-S Subhumid Glf Cstl Prair	4	0.040	--	0.040	0.050	0.068	--	0.090
34d-Coastal Sand Plain	0	--	--	--	--	--	--	--
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	0.040	--	--	--
34g-Texas-Louisiana Cstl Marsh	20	0.040	0.045	0.060	0.100	0.128	0.180	0.180
34h-Mid-Coast Barr Isl C Marsh	101	0.015	0.040	0.045	0.075	0.120	0.240	0.300
34i-Laguna Madre Br Isl C Mrs	16	0.040	0.040	0.040	0.046	0.072	0.135	0.180

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	7	10.9	10.9	11.0	12.5	31.1	34.2	36.9
34b-S Subhumid Glf Cstl Prair	4	4.04	--	4.26	4.49	4.65	--	4.66
34d-Coastal Sand Plain	0	--	--	--	--	--	--	--
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	4.95	--	--	--
34g-Texas-Louisiana Cstl Marsh	5	4.69	--	5.57	5.83	6.51	--	8.24
34h-Mid-Coast Barr Isl C Marsh	39	3.00	4.61	5.25	6.01	7.55	9.47	17.1
34i-Laguna Madre Br Isl C Mrs	16	3.00	3.00	3.51	7.86	14.8	17.8	18.8

Appendix 3.2. Classification and regression tree (CART) and non-parametric changepoint analysis code from geospatial analyses

ANALYSIS: TP VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TP ~ DEV + AG + DEVAG + FOR + WET, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=205 (652 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.09047248      0 1.0000000 1.0089319 0.2582039
2 0.04815967      1 0.9095275 0.9836883 0.2697916

Node number 1: 205 observations,    complexity param=0.09047248
  mean=0.145761, MSE=0.009393165
  left son=2 (187 obs) right son=3 (18 obs)
  Primary splits:
      DEV   < 0.418173   to the left,  improve=0.09047248, (0 missing)
      DEVAG < 0.33346    to the left,  improve=0.07105964, (0 missing)
      WET   < 0.0576105  to the left,  improve=0.06243525, (0 missing)
      FOR   < 0.04691736 to the left,  improve=0.05992079, (0 missing)
      AG    < 0.3621503  to the right, improve=0.02903751, (0 missing)

Node number 2: 187 observations
  mean=0.1367166, MSE=0.008573858

Node number 3: 18 observations
  mean=0.2397222, MSE=0.008226312
```

ANALYSIS: TP VS. %DEVELOPED (nCPA)

```
          cp        r2 mean left mean right pperm       5%      25%      50%      75%      95%
[1,] 0.378695 0.09047248 0.1367166  0.2397222 0.002 0.0658405 0.378695 0.378695 0.378695
0.5532925
```

ANALYSIS: TN VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = TN ~ DEV + AG + DEVAG + FOR + WET, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=159 (698 observations deleted due to missingness)
```

```
        CP nsplit rel error    xerror      xstd
1 0.12543144      0 1.0000000 1.0193133 0.1791164
2 0.04428305      1 0.8745686 0.9780269 0.1763173

Node number 1: 159 observations,    complexity param=0.1254314
  mean=0.9525472, MSE=0.08607889
  left son=2 (146 obs) right son=3 (13 obs)
  Primary splits:
      WET   < 0.0464365 to the right, improve=0.12543140, (0 missing)
      DEVAG < 0.3907505 to the left,  improve=0.08741474, (0 missing)
      AG    < 0.2504661 to the left,  improve=0.07455512, (0 missing)
      FOR   < 0.0658555 to the left,  improve=0.06640981, (0 missing)
      DEV   < 0.140019  to the right, improve=0.03797929, (0 missing)

Node number 2: 146 observations
  mean=0.9215411, MSE=0.0604478

Node number 3: 13 observations
  mean=1.300769, MSE=0.2418802
```

ANALYSIS: TN VS. %WETLAND (nCPA)

```
          cp       r2 mean left mean right pperm       5%      25%       50%       75%       95%

[1,] 0.0464365 0.1254314  1.300769  0.9215411 0.001 0.034718 0.038297 0.0464365 0.0464365
0.2108135
```

ANALYSIS: CHL-A SPEC VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = CHLASPEC ~ DEV + DEVAG + AG + FOR + WET, data = est,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=89 (972 observations deleted due to missingness)

        CP nsplit rel error   xerror      xstd
1 0.07165334      0 1.0000000 1.024355 0.2692747
2 0.02646778      1 0.9283467 1.072094 0.2347430

Node number 1: 89 observations,    complexity param=0.07165334
  mean=10.35815, MSE=15.87265
  left son=2 (69 obs) right son=3 (20 obs)
  Primary splits:
      DEV   < 0.1792775 to the left,  improve=0.07165334, (0 missing)
      DEVAG < 0.6110515 to the left,  improve=0.04924335, (0 missing)
```

```
    FOR   < 0.0711685 to the left,  improve=0.03583231, (0 missing)
    AG    < 0.0314075 to the left,  improve=0.02215091, (0 missing)
    WET   < 0.1719061 to the right, improve=0.01760316, (0 missing)

Node number 2: 69 observations,    complexity param=0.02646778
  mean=9.783986, MSE=6.55105
  left son=4 (25 obs) right son=5 (44 obs)
  Primary splits:
    DEV   < 0.1086436 to the right, improve=0.08271742, (0 missing)
    WET   < 0.0464365 to the right, improve=0.05815098, (0 missing)
    FOR   < 0.0726815 to the left,  improve=0.04521347, (0 missing)
    AG    < 0.4879205 to the left,  improve=0.03784116, (0 missing)
    DEVAG < 0.185038  to the left,  improve=0.03237564, (0 missing)

Node number 3: 20 observations
  mean=12.339, MSE=42.97105

Node number 4: 25 observations
  mean=8.8074, MSE=9.604186

Node number 5: 44 observations
  mean=10.33886, MSE=3.966537
```

ANALYSIS: CHL-A SPEC VS. %DEVELOPED (nCPA)

```
         cp        r2 mean left mean right pperm       5%       25%       50%      75%      95%

[1,] 0.1792775 0.07165334  9.783986     12.339  0.13 0.1048592 0.1587125 0.1792775 0.378695
0.667554
```

ANALYSIS: CHL-A FLUORO VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ DEV + DEVAG + AG + FOR + WET, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=71 (990 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.18975008      0   1.00000 1.028180 0.3733131
2 0.04161515      1   0.81025 1.042686 0.3470457

Node number 1: 71 observations,    complexity param=0.1897501
  mean=8.437324, MSE=41.4621
  left son=2 (58 obs) right son=3 (13 obs)
```

```
  Primary splits:
      DEVAG < 0.6110515 to the left,  improve=0.18975010, (0 missing)
      AG    < 0.506128  to the left,  improve=0.07381704, (0 missing)
      FOR   < 0.051275  to the left,  improve=0.05517455, (0 missing)
      WET   < 0.0464365 to the right, improve=0.04746868, (0 missing)
      DEV   < 0.1587125 to the left,  improve=0.03740028, (0 missing)

Node number 2: 58 observations
  mean=7.109397, MSE=14.9794

Node number 3: 13 observations
  mean=14.36192, MSE=116.6474
```

ANALYSIS: CHL-A FLUORO VS. DEVELOPED+AGRICULTURE (nCPA)

```
          cp        r2 mean left mean right pperm       5%       25%      50%       75%      95%

[1,] 0.6110515 0.1897501  7.109397   14.36192 0.003 0.185038 0.5212905 0.602696 0.6110515
0.648743
```

ANALYSIS: SECCHI TRANSPARENCY VS. WATERSHED LULC (CART)

```
Call:
mvpart(form = SECCHI ~ DEV + DEVAG + AG + FOR + WET, data = est,
    xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=223 (838 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.05685374      0 1.0000000 1.005813 0.1621427
2 0.04101964      1 0.9431463 1.044457 0.1620465

Node number 1: 223 observations,    complexity param=0.05685374
  mean=0.6052915, MSE=0.04994509
  left son=2 (83 obs) right son=3 (140 obs)
  Primary splits:
      DEVAG < 0.4206727 to the right, improve=0.05685374, (0 missing)
      AG    < 0.269731  to the right, improve=0.04605632, (0 missing)
      DEV   < 0.094233  to the left,  improve=0.03673752, (0 missing)
      FOR   < 0.1697195 to the left,  improve=0.03184019, (0 missing)
      WET   < 0.0576105 to the right, improve=0.02677017, (0 missing)

Node number 2: 83 observations
  mean=0.5360843, MSE=0.03061629
```

```
Node number 3: 140 observations
  mean=0.6463214, MSE=0.05688129
```

ANALYSIS: SECCHI TRANSPARENCY VS. %DEVELOPED+AGRICULTURE

```
        cp       r2 mean left mean right pperm      5%     25%      50%      75%      95%

[1,] 0.4206727 0.05685374 0.6463214  0.5360843 0.006 0.185038 0.185038 0.3409165 0.4206727
0.5975985
```

ANALYSIS: TP VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (CART)

```
Call:
mvpart(form = TP ~ FLOW, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=205 (652 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.05552219      0 1.0000000 1.0090249 0.2555688
2 0.03297608      3 0.8298364 0.8820702 0.2302095

Node number 1: 205 observations,    complexity param=0.05552219
  mean=0.145761, MSE=0.009393165
  left son=2 (139 obs) right son=3 (66 obs)
  Primary splits:
      FLOW < 622.7558 to the left,  improve=0.04024549, (0 missing)

Node number 2: 139 observations,    complexity param=0.05552219
  mean=0.1323633, MSE=0.009839855
  left son=4 (75 obs) right son=5 (64 obs)
  Primary splits:
      FLOW < 10.89075 to the right, improve=0.09967557, (0 missing)

Node number 3: 66 observations
  mean=0.1739773, MSE=0.007278215

Node number 4: 75 observations
  mean=0.1034333, MSE=0.003465882

Node number 5: 64 observations,    complexity param=0.05552219
  mean=0.1662656, MSE=0.0151792
  left son=10 (30 obs) right son=11 (34 obs)
  Primary splits:
```

```
    FLOW < 6.1195   to the left,  improve=0.1171834, (0 missing)

Node number 10: 30 observations
  mean=0.1213667, MSE=0.004686832

Node number 11: 34 observations
  mean=0.2058824, MSE=0.02108893
```

ANALYSIS: TP VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED (nCPA)

```
     cp        r2 mean left mean right pperm      5%      25%      50%

[1,] 622.7558 0.04024549 0.1323633  0.1739773 0.033 8.37529 10.89075 463.4954

         75%      95%

[1,] 622.7558 622.7558
```

ANALYSIS: TP VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE WEIGHTED (CART)

```
Call:
mvpart(form = TP ~ WFLOW, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=205 (856 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.04699010      0 1.0000000 1.009156 0.2561318
2 0.03523812      2 0.9060198 1.029260 0.2746812

Node number 1: 205 observations,    complexity param=0.0469901
  mean=0.145761, MSE=0.009393165
  left son=2 (132 obs) right son=3 (73 obs)
  Primary splits:
      WFLOW < 0.1124294   to the left,  improve=0.04674533, (0 missing)

Node number 2: 132 observations,    complexity param=0.0469901
  mean=0.130178, MSE=0.009897629
  left son=4 (72 obs) right son=5 (60 obs)
  Primary splits:
      WFLOW < 0.005140514 to the right, improve=0.06961831, (0 missing)

Node number 3: 73 observations
```

```
  mean=0.1739384, MSE=0.007247931

Node number 4: 72 observations
  mean=0.1062153, MSE=0.00493811

Node number 5: 60 observations
  mean=0.1589333, MSE=0.01433313
```

ANALYSIS: TP VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE WEIGHTED (nCPA)

```
          cp        r2 mean left mean right pperm        5%       25%

[1,] 0.1124294 0.04674533 0.1301780  0.1739384 0.029 0.003375662 0.1124294

         50%       75%       95%

[1,] 0.1124294 0.1124294 0.2410194
```

ANALYSIS: TN VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (CART)

```
Call:
mvpart(form = TN ~ FLOW, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=159 (698 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.02561440      0 1.0000000 1.004226 0.1770573
2 0.02521484      1 0.9743856 1.103708 0.1942988

Node number 1: 159 observations,    complexity param=0.0256144
  mean=0.9525472, MSE=0.08607889
  left son=2 (107 obs) right son=3 (52 obs)
  Primary splits:
      FLOW < 11.62225 to the right, improve=0.0256144, (0 missing)

Node number 2: 107 observations
  mean=0.9198131, MSE=0.07329249

Node number 3: 52 observations
  mean=1.019904, MSE=0.1056476
```

ANALYSIS: TN VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED (nCPA)

```
      cp       r2 mean left mean right pperm    5%      25%    50%

[1,] 11.62225 0.0256144  1.019904   0.919813 0.257 6.405 11.62225 14.761

      75%      95%

[1,] 466.1183 949.8036
```

ANALYSIS: TN VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE WEIGHTED (CART)

```
Call:
mvpart(form = TN ~ WFLOW, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=159 (902 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.04413561      0 1.0000000 1.008476 0.1778643
2 0.01548700      1 0.9558644 1.011164 0.1693175

Node number 1: 159 observations,    complexity param=0.04413561
  mean=0.9525472, MSE=0.08607889
  left son=2 (102 obs) right son=3 (57 obs)
  Primary splits:
      WFLOW < 0.007352551 to the right, improve=0.04413561, (0 missing)

Node number 2: 102 observations
  mean=0.9064706, MSE=0.05726548

Node number 3: 57 observations
  mean=1.035, MSE=0.1270421
```

ANALYSIS: TN VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE WEIGHTED (nCPA)

```
          cp         r2 mean left mean right pperm          5%         25%

[1,] 0.007352551 0.04413561     1.035  0.9064706 0.063 0.001689401 0.007352551

             50%        75%      95%

[1,] 0.007352551 0.01043468 1.186907
```

ANALYSIS: CHL-A SPEC VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (CART)

```
Call:
mvpart(form = CHLASPEC ~ FLOW + WFLOW, data = est, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
  n=89 (972 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.02925510      0  1.000000 1.017266 0.2697765
2 0.01352799      1  0.970745 1.149087 0.2802860

Node number 1: 89 observations,    complexity param=0.0292551
  mean=10.35815, MSE=15.87265
  left son=2 (78 obs) right son=3 (11 obs)
  Primary splits:
      FLOW  < 622.7558  to the left,  improve=0.0292551, (0 missing)
      WFLOW < 0.2410194 to the left,  improve=0.0292551, (0 missing)

Node number 2: 78 observations
  mean=10.10224, MSE=10.74251

Node number 3: 11 observations
  mean=12.17273, MSE=48.49289
```

ANALYSIS: CHL-A SPEC VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (nCPA)

```
           cp         r2 mean left mean right pperm     5%    25%     50%      75%      95%

[1,] 442.0598 0.02619934  10.10224   12.17273 0.519 2.2865 14.761 100.037 442.0598 622.7558

            cp         r2 mean left mean right pperm          5%         25%        50%         75%         95%

[1,] 0.1342000 0.02140125  10.10224   12.17273 0.513 0.001689401 0.04612783 0.0827481 0.2040619
0.2410194
```

ANALYSIS: CHL-A FLUORO VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ FLOW + WFLOW, data = est, xval = 10,
   method = "anova", minsplit = 20, minbucket = 10)
  n=71 (990 observations deleted due to missingness)
```

```
        CP nsplit rel error   xerror      xstd
1 0.09254536      0 1.0000000 1.051278 0.3728333
2 0.01835181      1 0.9074546 1.108913 0.3619334

Node number 1: 71 observations,    complexity param=0.09254536
  mean=8.437324, MSE=41.4621
  left son=2 (54 obs) right son=3 (17 obs)
  Primary splits:
      FLOW  < 8.37529    to the right, improve=0.09254536, (0 missing)
      WFLOW < 0.00223445 to the right, improve=0.04056874, (0 missing)

Node number 2: 54 observations
  mean=7.338241, MSE=18.02642

Node number 3: 17 observations
  mean=11.92853, MSE=99.87923
```

ANALYSIS: CHL-A SPEC VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (nCPA)

```
        cp       r2 mean left mean right pperm    5%     25%     50%      75%    95%

[1,] 8.37529 0.09254536  11.92853   7.338241 0.097 7.2155 8.37529 8.37529 11.62225 76.535

            cp        r2 mean left mean right pperm         5%        25%        50%        75%        95%

[1,] 0.002234450 0.04056874  10.74882    7.70963  0.38 0.001439732 0.002234450 0.007352551
0.07643062 0.07996369
```

ANALYSIS: SECCHI TRANSPARENCY VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (CART)

```
Call:
mvpart(form = SECCHI ~ FLOW + WFLOW, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=223 (838 observations deleted due to missingness)

        CP nsplit rel error    xerror      xstd
1 0.04949755      0 1.0000000 1.0069318 0.1633173
2 0.01413505      2 0.9010049 0.9613792 0.1688737

Node number 1: 223 observations,    complexity param=0.04949755
  mean=0.6052915, MSE=0.04994509
  left son=2 (68 obs) right son=3 (154 obs), 1 observation remains
  Primary splits:
```

```
    WFLOW < 0.2410194   to the right, improve=0.03449432, (1 missing)
    FLOW  < 622.7558    to the right, improve=0.03390651, (0 missing)

Node number 2: 68 observations
  mean=0.5431618, MSE=0.01223596

Node number 3: 154 observations,    complexity param=0.04949755
  mean=0.6334091, MSE=0.06435315
  left son=6 (81 obs) right son=7 (73 obs)
  Primary splits:
    FLOW  < 13.66169    to the left,  improve=0.07136531, (0 missing)
    WFLOW < 0.005140513 to the left,  improve=0.06560521, (0 missing)

Node number 6: 81 observations
  mean=0.5690741, MSE=0.04893525

Node number 7: 73 observations
  mean=0.7047945, MSE=0.07177222
```

ANALYSIS: SECCHI TRANSPARENCY VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT DISCHARGE UNWEIGHTED AND WEIGHTED (nCPA)

```
         cp        r2 mean left mean right pperm      5%     25%      50%      75%      95%
[1,] 622.7558 0.03390651 0.6325484  0.5431618 0.059 0.899275 12.0559 463.4954 622.7558 622.7558
          cp        r2 mean left mean right pperm         5%        25%       50%       75%       95%
[1,] 0.2410194 0.03452885 0.6334091  0.5431618 0.077 0.000291451 0.005140513 0.2410194
0.2410194 0.2410194
```

ANALYSIS: TP VS. SALINITY (CART)

```
Call:
mvpart(form = TP ~ SALINITY, data = est, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=206 (855 observations deleted due to missingness)

       CP nsplit rel error    xerror      xstd
1 0.3160122      0 1.0000000 1.0067412 0.2541140
2 0.0234496      1 0.6839878 0.7676378 0.2571038

Node number 1: 206 observations,    complexity param=0.3160122
  mean=0.1453447, MSE=0.009383097
  left son=2 (111 obs) right son=3 (95 obs)
  Primary splits:
    SALINITY < 17.43 to the right, improve=0.3160122, (0 missing)

Node number 2: 111 observations
```

```
  mean=0.09496847, MSE=0.001467199

Node number 3: 95 observations
  mean=0.2042053, MSE=0.01220246
```

ANALYSIS: TP VS. SALINITY (nCPA)

cp	r2	mean left	mean right	pperm
17.43	0.3160122	0.2042053	0.09496847	0.001
5%	25%	50%	75%	95%
16.35	16.94083	17.43	18.26542	19.32146

ANALYSIS: TN VS. SALINITY (CART)

```
Call:
mvpart(form = TN ~ SALINITY, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=160 (901 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.06257217      0 1.0000000 1.009700 0.1789083
2 0.06063811      1 0.9374278 1.089849 0.1947610

Node number 1: 160 observations,    complexity param=0.06257217
  mean=0.9514687, MSE=0.08572581
  left son=2 (98 obs) right son=3 (62 obs)
  Primary splits:
      SALINITY < 17.1375 to the right, improve=0.06257217, (0 missing)

Node number 2: 98 observations
  mean=0.8932143, MSE=0.08486441

Node number 3: 62 observations
  mean=1.043548, MSE=0.07324467
```

ANALYSIS: TN VS. SALINITY (nCPA)

cp	r2	mean left	mean right	pperm
17.1375	0.06257217	1.043548	0.8932143	0.028
5%	25%	50%	75%	95%
12.8075	16.94083	20.06667	21.29938	31.82542

ANALYSIS: TN VS. SALINITY AFTER REMOVING OUTLIERS (CART)

```
Call:
mvpart(form = TN ~ SAL, data = est, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=154 (703 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.20302623      0 1.0000000 1.0196281 0.2011627
2 0.01994302      1 0.7969738 0.8464945 0.1644317

Node number 1: 154 observations,    complexity param=0.2030262
  mean=0.9192208, MSE=0.06021011
  left son=2 (65 obs) right son=3 (89 obs)
  Primary splits:
      SAL < 21.19937 to the right, improve=0.2030262, (0 missing)

Node number 2: 65 observations
  mean=0.7898462, MSE=0.03355382

Node number 3: 89 observations
  mean=1.013708, MSE=0.05852614
```

ANALYSIS: TN VS. SALINITY AFTER REMOVING OUTLIERS (nCPA)

```
           cp        r2 mean left mean right pperm    5%      25%    50%      75%      95%

[1,] 21.19937 0.2030262  1.013708  0.7898462 0.001 16.95 19.94479 20.975 21.19937 21.59938
```

ANALYSIS: CHL-A SPEC VS. SALINITY (CART)

```
Call:
mvpart(form = CHLASPEC ~ SALINITY, data = est, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=90 (971 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.07760569      0 1.0000000 1.053564 0.2781870
2 0.06100064      1 0.9223943 1.106067 0.2897142

Node number 1: 90 observations,    complexity param=0.07760569
  mean=10.35417, MSE=15.69769
  left son=2 (80 obs) right son=3 (10 obs)
  Primary splits:
      SALINITY < 10.53 to the right, improve=0.07760569, (0 missing)
```

```
Node number 2: 80 observations
  mean=9.963938, MSE=11.14449

Node number 3: 10 observations
  mean=13.476, MSE=41.15928
```

ANALYSIS: CHL-A SPEC VS. SALINITY (nCPA)

```
       cp         r2 mean left mean right pperm       5%    25%    50%     75%   95%
[1,] 10.5 0.08765997     13.476    9.963938 0.058 9.801667 10.25 11.85 25.325 33.8
```

ANALYSIS: CHL-A FLUORO VS. SALINITY (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ SALINITY, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=72 (989 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.08322143      0 1.0000000 1.023223 0.3630847
2 0.02465243      1 0.9167786 1.106407 0.3616209

Node number 1: 72 observations,    complexity param=0.08322143
  mean=8.497917, MSE=41.14692
  left son=2 (54 obs) right son=3 (18 obs)
  Primary splits:
      SALINITY < 17.23667 to the right, improve=0.08322143, (0 missing)

Node number 2: 54 observations
  mean=7.429537, MSE=17.55709

Node number 3: 18 observations
  mean=11.70306, MSE=98.21919
```

ANALYSIS: CHL-A FLUORO VS. SALINITY (nCPA)

```
           cp         r2 mean left mean right pperm    5%      25%      50%    75%     95%
[1,] 17.23667 0.08322143  11.70306   7.429537 0.101 10.56 14.76667 17.23667 19.825 33.8375
```

ANALYSIS: SECCHI TRANSPARENCY VS. SALINITY (CART)

```
Call:
mvpart(form = SECCHI ~ SALINITY, data = est, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
```

```
  n=224 (837 observations deleted due to missingness)

        CP nsplit rel error    xerror      xstd
1 0.33737153      0 1.0000000 1.0104597 0.1628001
2 0.03838564      1 0.6626285 0.7228622 0.1204500

Node number 1: 224 observations,    complexity param=0.3373715
  mean=0.6062054, MSE=0.04990837
  left son=2 (124 obs) right son=3 (100 obs)
  Primary splits:
      SALINITY < 20.82437 to the left,  improve=0.3373715, (0 missing)

Node number 2: 124 observations
  mean=0.4896774, MSE=0.01018094

Node number 3: 100 observations
  mean=0.7507, MSE=0.06145401
```

ANALYSIS: SECCHI TRANSPARENCY VS. SALINITY(nCPA)

```
         cp       r2 mean left mean right pperm   5%      25%  50%      75%      95%

[1,] 20.82437 0.3373715 0.4896774     0.7507 0.001 20.6 20.82437 21.5 21.70833 22.66667
```

Appendix 3.3. Classification and regression tree (CART) and non-parametric changepoint analysis code from stressor response analyses.

ANALYSIS: SECCHI TRANSPARENCY VS. NUTRIENTS (TP, TN , NOX-N, NH4-N, SRP) (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + NOX + NH4 + SRP, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=208 (649 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.3089278      0 1.0000000 1.0103795 0.1777142
2 0.1574233      1 0.6910722 0.8487537 0.1249873

Node number 1: 208 observations,    complexity param=0.3089278
  mean=0.5892548, MSE=0.04506663
  left son=2 (171 obs) right son=3 (33 obs), 4 observations remain
  Primary splits:
      TP   < 0.0675 to the right, improve=0.28483730, (4 missing)
      TN   < 0.845  to the right, improve=0.09053712, (50 missing)
      SRP  < 0.0625 to the right, improve=0.07840979, (9 missing)
      NOX  < 0.0425 to the right, improve=0.04978828, (8 missing)

Node number 2: 171 observations,    complexity param=0.1574233
  mean=0.5362865, MSE=0.02498811
  left son=4 (97 obs) right son=5 (53 obs), 21 observations remain
  Primary splits:
      TP   < 0.1125 to the right, improve=0.13404710, (0 missing)
      NH4  < 0.0975 to the right, improve=0.03804167, (4 missing)
      TN   < 0.845  to the right, improve=0.02360814, (41 missing)
      NOX  < 0.5275 to the right, improve=0.01990184, (8 missing)

Node number 3: 33 observations
  mean=0.8469697, MSE=0.06681961

Node number 4: 97 observations
  mean=0.4970103, MSE=0.00842199

Node number 5: 53 observations
  mean=0.6392453, MSE=0.03736547
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP (nCPA)

```
       cp       r2 mean left mean right pperm    5%    25%    50%    75%
[1,] 0.0675 0.2918685 0.8469697  0.5362865 0.001 0.0625 0.0625 0.0675 0.0755
      95%
[1,] 0.105
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN (nCPA)

```
       cp       r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 0.845 0.1367112 0.7137037  0.5591827 0.001 0.635 0.695 0.735 0.845 0.855
```

ANALYSIS: CHL-A SPEC VS. NUTRIENTS (TP, TN , NOX-N, NH4-N, SRP) (CART)

```
Call:
mvpart(form = CHLASPEC ~ TN + TP + SRP + NOX + NH4, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=90 (767 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.15012793      0  1.000000 1.014390 0.2690103
2 0.06950896      1  0.849872 1.023839 0.2459605

Node number 1: 90 observations,    complexity param=0.1501279
  mean=10.35417, MSE=15.69769
  left son=2 (79 obs) right son=3 (11 obs)
  Primary splits:
     TP  < 0.2775 to the left,  improve=0.15012790, (0 missing)
     TN  < 1.42   to the left,  improve=0.12201810, (25 missing)
     SRP < 0.135  to the left,  improve=0.07390838, (1 missing)
     NOX < 0.0675 to the left,  improve=0.04411884, (0 missing)

Node number 2: 79 observations
  mean=9.781329, MSE=9.818626

Node number 3: 11 observations
  mean=14.46818, MSE=38.63831Call:
mvpart(form = CHLASPEC ~ TN + TP + SRP + NOX + NH4 + SAL, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=90 (767 observations deleted due to missingness)
```

ANALYSIS: CHL-A SPEC VS. TP (nCPA)

```
      cp       r2 mean left mean right pperm    5%   25%   50%  75%  95%

[1,] 0.25 0.1258186  9.931494   12.85769 0.004 0.1375 0.151 0.205 0.27 0.28
```

ANALYSIS: CHL-A SPEC VS. TN (nCPA)

```
       cp       r2 mean left mean right pperm    5%  25%   50%    75%  95%

[1,] 1.265 0.4389467  10.05741   14.15455 0.001 1.105 1.24 1.265 1.3875 1.55
```

ANALYSIS: CHL-A FLUORO VS. NUTRIENTS (TP, TN , NOX-N, NH4-N, SRP) (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TN + TP + SRP + NOX + NH4 + SAL, data = est,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=72 (785 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.49898230      0 1.0000000 1.032438 0.3659925
2 0.05019523      1 0.5010177 0.928954 0.2182731

Node number 1: 72 observations,    complexity param=0.4989823
  mean=8.497917, MSE=41.14692
  left son=2 (52 obs) right son=3 (13 obs), 7 observations remain
  Primary splits:
     TN  < 1.115    to the left,  improve=0.44863170, (7 missing)
     SRP < 0.1075   to the left,  improve=0.28474580, (0 missing)
     TP  < 0.1075   to the left,  improve=0.25234800, (0 missing)
     NOX < 0.1375   to the left,  improve=0.16669990, (0 missing)
     SAL < 17.23667 to the right, improve=0.08322143, (0 missing)

Node number 2: 52 observations
  mean=6.418269, MSE=6.415636
Node number 3: 13 observations  mean=17.72308, MSE=88.51468
```

ANALYSIS: CHL-A FLUORO VS. TP (nCPA)

```
       cp       r2 mean left mean right pperm    5%   25%   50%  75%   95%

[1,] 0.1075 0.2523480  6.499519   13.69375 0.002 0.085 0.105 0.115 0.13 0.151
```

ANALYSIS: CHL-A FLUORO VS. TN (nCPA)

	cp	r2	mean left	mean right	pperm	5%	25%	50%	75%	95%
[1,]	1.115	0.4724182	6.418269	17.72308	0.001	0.96	1.09	1.1125	1.1825	1.375

Section 4: Tidal Streams

E.M. Grantz, L.B. Massey, B.E. Haggard, and J.T. Scott

EXECUTIVE SUMMARY

The Clean Water Action Plan, released in 1998 by the United States Environmental Protection Agency (USEPA), established a national set of nutrient criteria for the 14 aggregate ecoregions across the United States, directing states and tribes to adopt these criteria or pursue scientifically defensible nutrient criteria at the state level. For streams and rivers, including tidal streams, the main approaches for nutrient criteria development focus on the frequency distribution of median concentrations of a general population or select group of sites representing reference conditions and statistical analysis of stressor-response relationships between nutrients and biological response variables.

The objective of Section 4 was to provide statistical support to the Texas Commission on Environmental Quality (TCEQ) to aid the development of numeric nutrient criteria for Texas tidal streams. First, geospatial and water quality data from 254 stations spanning 18 basins were compiled. These data were provided by TCEQ and collected under non-biased conditions. Following data reorganization and reduction, medians were calculated for each parameter and station where 10 observations or greater were available and compiled into a database. The parameters of concern were total phosphorus (TP), ortho-phosphate (PO_4-P; SRP), total nitrogen (TN), nitrate plus nitrite N (NO_x-N), and sestonic chlorophyll-a (chl-a). Frequency distributions were calculated at multiple spatial scales, specifically, by basin, eco-region levels III and IV, and basin by level III eco-region and are presented in Section 4.1.

Section 4.2 provides analyses of potential geospatial variability in total nutrient (TP and TN), chl-a, and Secchi transparency for Texas tidal streams using classification and regression tree (CART) and non-parametric changepoint analysis (nCPA). Geospatial variables included land use/land cover (LULC) categories, permitted municipal wastewater treatment (WWTP) plant discharge, and basin. In contrast to the general streams populations, LULC variables were good predictors of nutrient and biological parameters in tidal streams. For both TN and TP, 8-9% wetland land cover was identified as a threshold, while 85% developed+agricultural and 29% agricultural were thresholds for TN and Secchi transparency, respectively. Thresholds in WWTP discharge were found for TP and were in range with those found for the larger streams and rivers database (~4-5 mgd). All parameters of interest were grouped by basin, with low and high groups accounting for 11-36% of variability in the data.

Section 4.3 provides analyses of potential nutrient thresholds (TP, TN) to biological response (chl-a measured spectrophotometrically and fluorometrically, Secchi transparency) in Texas tidal streams. The only statistically significant relationship supported by ecological theory found was a TP threshold in Secchi transparency of 0.088 mg/L.

INTRODUCTION

The Clean Water Action Plan released in 1998, established a national set a nutrient criteria for the 14 aggregate eco-regions across the United States (US), five of which lie partly within Texas. These numerical values were set for both stressor (e.g., nutrients) and response (e.g., chlorophyll and transparency) variables, based on frequency distributions. However, local and regional influences on water quality can contribute to median concentrations that are different than what the US Environmental Protection Agency (USEPA) has recommended (e.g., Ice et al. 2003, Smith et al. 2003, Binkley 2004, Evans-White et al. 2013). For example, slight differences were observed in the 25th percentiles for the Red River Basin flowing from New Mexico to Louisiana compared to that recommended by the USEPA (Longing and Haggard 2010). The aggregate eco-regions may be too coarse to be used for establishing nutrient criteria, and the basin or smaller eco-region level might be more appropriate for the development of nutrient criteria (Rohm et al. 2002). Therefore, states, tribes and others have the option of adopting the criteria set by the USEPA or establishing scientifically defensive nutrient criteria for water bodies of various spatial scales (e.g., basins and ecoregions) specific to a regulatory jurisdiction.

There are two commonly applied methods to evaluating nutrient concentrations in flowing waters for the development of numeric criteria, including frequency distributions and stressor-response studies. The frequency distribution method develops nutrient criteria relative to the population of water-quality data in a specific area (e.g., state, basin or ecoregion). The USEPA (2000) has suggested the 75th percentile of nutrient concentrations from reference or minimally impacted streams conditions as a criterion, or the 25th percentile of nutrient concentrations from a general population (i.e., all streams regardless of human influence). The USEPA (2000) suggests that both approaches should result in similar criterion; however, studies have shown that a comparison between approaches can be highly variable (Suplee et al. 2007, Herlihy and Sifeneos 2008). There are many concerns with this approach, such as limited data representing reference or even general populations from targeted areas and the selected percentile is not necessarily tied to water-quality impairments. Nonetheless, the frequency distribution method is a tool that can aid states, tribes and other groups when setting nutrient criteria.

The USEPA has recommended that states and tribes use stressor-response studies to help develop nutrient criteria, where biological conditions are evaluated over a gradient of nutrient concentrations. Classification and regression tree (CART) analysis is an empirical modeling technique that is useful for identifying ecological thresholds and hierarchical structure in predictor variables (De'ath and Fabricius 2000). CART uses recursive partitioning to divide data into subsets that are increasingly homogeneous, invoking a tree-like classification that can explain relationships that may be difficult to reconcile with conventional linear models (Urban 2002). CART and other similar methods have been used to identify thresholds and hierarchical structure in environmental correlates of various biological processes in aquatic ecosystems (King et al. 2005, East and Sharfstein 2006). King et al. (2005) used CART to identify thresholds in nutrient concentrations which resulted in shifts in ecological structure and function. These thresholds were used to recommend specific water quality nutrient criteria for the Florida Everglades ecosystem. A review article has been recently released providing a comprehensive review of nutrient thresholds identified through various stressor-response studies, as well as comparison of frequency distributions across the US (Evans-White et al. 2013).

States across the US are moving forward with the development of numeric nutrient criteria, although the pace varies by state and the political, legal and environmental pressures each state is facing. Many states are concerned about the legitimacy of promulgating one numeric criterion across the whole state comprised by multiple basins, various level III and IV ecoregions, and different land uses (e.g., forest, pasture, row crop and urban). However, the development of site-specific nutrient criteria can be a costly process from the efforts needed to evaluate the physical, chemical and biological conditions of flowing waters to that required to push the numeric criteria through promulgation. For these reasons, it might not be feasible to develop numeric criteria for individual watersheds or eco-regions. However, studies have shown that almost half of the variation in nutrient concentrations can be explained by select physico-chemical properties and watershed characteristics, like runoff, elevation, land use and cover, and also eco-regions (Sifneos and Herlihy 2008). There is also evidence to suggest that undisturbed watershed conditions may not exist because of the effects of even minimal development (King and Baker 2010), atmospheric deposition (Flum and Nodvin 1995) and small catchment areas (Smith et al. 2003). Thus, states need to explore defensible approaches to aggregating stream stations into categories to assist in the nutrient criteria development process.

The objectives of this section are:

1) to discuss the frequency distribution of median nutrient concentrations and response variables for Texas tidal streams acquired from the Texas Commission on Environmental Quality (TCEQ) at various spatial scales including individual basins, level III ecoregions, level IV ecoregions, and basin-level III ecoregion combinations;
2) to explore the relationship between median nutrient concentrations (focusing on TP and TN), as well as common biological parameters (focusing on Secchi transparency and chlorophyll-a) and watershed attributes (both numeric and categorical) for Texas tidal streams, providing a defensible approach from which Texas tidal streams could be grouped by watershed attributes;
3) to identify nutrient threshold values associated with changes in the magnitude or variability of commonly measured biological response variables for Texas tidal stream.

The approach used for the tidal stream data replicate that used for the other water body types, and it is important to note that there was little information in the literature distinguishing tidal streams from other freshwater streams flowing across the landscape.

4.1: TIDAL STREAMS DATABASE DEVELOPMENT, MEDIAN CALCULATION, AND FREQUENCY DISTRIBUTIONS

Methods

Water Quality Database

Data Acquisition, Compilation and Reduction. TCEQ provided a database of water quality data collected from 1968 to 2012 from tidal streams along the Texas coastline. The collected data was from 254 stations spanning 18 watersheds. The data described 116 tidal stream characteristics and water quality parameters including nutrients, sediments, transparency, physico-chemical parameters, as well as others.

For the purposes of advanced statistical analyses conducted during this project, only data collected under specific monitoring type codes (as decided by TCEQ) and from 2000 to 2010 was used. Therefore, the database was sorted and any data collected before calendar year 2000 or after 2010 was removed. Data collected under the monitoring type code Biased Flow (BF) was also removed since data collected under this circumstance were not necessarily representative of baseline water quality conditions. The data received from TCEQ were output to a single column format within the files, so the data were reorganized into a useable format. The data was sorted by Basin ID and a new Microsoft Excel worksheet was created for each individual basin. Each basin worksheet was then restructured using the pivot table function in Microsoft Excel so that each parameter and the associated data were unique to an individual column; a portion of this process was accomplished with a Mircosoft Excel Macro (see Appendix 1.1 for Excel Macro code). Any estimated data points (i.e., those reported with a < or >) were flagged and used in the database without the associated qualifying sign. The data were flagged using a Microsoft Excel Macro (Appendix 1.2).

Several additional parameters were calculated from the original data provided. Nitrate plus nitrite and total N (TN) were calculated if the necessary N species were provided by TCEQ in the original data file. In addition, diel change (i.e., 24 hour maximum minus 24 hour minimum) was calculated for dissolved oxygen, temperature, conductivity, pH, and turbidity. The additional parameters were added to each station worksheet using a Microsoft Excel Macro (Appendix 1.3).

Due to the volume of data provided, several parameters were removed from the median database because of lack of data and duplication of parameters, or because TCEQ indicated that the parameter could be removed from the database.

Median and Frequency Distribution Calculations. For this study, frequency distribution and, subsequently, stressor-response analyses were conducted on station medians in order to focus on broadly applicable regional and statewide trends. Because each tidal stream in Texas was not equally represented in the raw water quality dataset, conducting statistical analyses on medians removes potential site-specific bias for sites that are over- or under-represented in the raw dataset. Furthermore, biological response and nutrient stressor data did not always overlap in the raw data. Median values of each parameter were calculated for each Station ID using a Microsoft Excel Macro (Appendix 1.4). Median values were calculated based on at least 10 data points, i.e. no medians were calculated if less than 10 data points were available for a given parameter at a given station. The calculated medians for each Station ID were then compiled into one database using a Microsoft Excel Macro (Appendix 1.5). This database was merged with the GIS and LULC data and used in advanced statistical analysis.

Frequency distributions (minimum value, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles and maximum value) for water quality parameters TP (TCEQ parameter code 00665), TN (calculated parameter code 00600C; TCEQ parameter code 00625 + 00630, 00625 + 00593 or 00625 + 00615 + 00620), NO_x-N (calculated parameter 00630C; TCEQ parameter code 00630, 00593 or 00615 + 00620), PO_4-P (TCEQ calculated parameter code 00671C; TCEQ parameter code 00671 or 70507), and chl-a (TCEQ parameter code 70953) were calculated using Microsoft Excel. For this study, a parameter combining chl-a measured spectrophotometrically (parameter code 32211; chl-a spec) and chl-a fluoro was not created due to inconsistencies between the methods (Laurie Eng, personal communication). Data were more complete and censorship was less of a concern for chl-a fluoro than for chl-a spec. Spectrophotometric chl-a data were commonly censored at a relatively high detection limit (10 µg/L). Analysis exploring the effects of censored data on chl-a spec median calculation in Texas reservoirs indicated that, when censored data exceeded 16% of the raw data for a station, chl-a medians (i.e. the 50^{th} percentile) were increasingly overestimated when the detection limit was substituted for censored observations as the level of censoring increased because the median could not be calculated to be a value below the detection limit (see Section 2.3). This censored data effect would have been magnified further when considering very low percentiles in the frequency distribution, such as the 25^{th} percentile. Therefore, frequency distributions for sestonic chl-a were only calculated for the fluorometric method in this study.

Geospatial Database

A geospatial database contained within a Microsoft Excel file was provided by TCEQ that identified land use and land cover data for the water quality stations located within tidal streams included in this study. The geospatial descriptors were provided for the drainage basin, with the upstream boundary constrained by the nearest upstream reservoir. The descriptors included percent open water, developed-open, developed-low intensity, developed- medium intensity, developed-high intensity, barren land, deciduous forest, evergreen forest, mixed forest, shrub/scrubland, grassland/herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands. These descriptors were reduced to five categories including percent developed (i.e., open, low intensity, medium intensity, barren land), forest (i.e., deciduous, evergreen, mixed and shrubland), agriculture (i.e., grassland/herbaceous, pasture/hay, cultivated crops), developed plus agriculture (i.e., open, low intensity, medium intensity, barren land, grassland/herbaceous, pasture/hay, cultivated crops) and wetlands (i.e., woody and emergent herbaceous wetlands). Additional geospatial information for each site was provided including drainage area, slope, municipal discharges, basin ID, level III ecoregion ID, and level IV ecoregion ID.

Data Quality Assurance and Control

Data quality checks were employed frequently throughout the database reorganization and data calculation processes. The original source files were maintained in an unaltered form, and subsequent changes to each database were saved under unique file names. Data transferred from one file to the next were checked for accuracy by comparing first and last rows and the row count between files. In addition, when calculations were preformed, including manual calculations and those calculated using Microsoft Excel Macros, at least 10 percent of calculations were checked for accuracy following the secondary data quality assurance project plan (QAPP).

Results and Discussion

Basin. The State of Texas is divided into 23 basins (Appendix 1.7) which are categorized as river (65%) or coastal (35%) basin waters. River basin waters are the surface inland waters comprising the major streams and their tributaries while coastal basin waters are surface inland waters that discharge or in some way interconnect with bays or the Gulf of Mexico. Texas streams classified as tidal are located within 18 of these basins. Of these basins, sufficient TP medians were available to estimate the 25th percentile of median TP concentrations for 8 of these basins, and, for 75% of these basins, 25th percentiles exceeded 0.10 mg/L, ranging from 0.070 to 0.41 mg/L (Table 4.1.1). For these same basins, sufficient data was available to estimate 25th percentiles for PO_4-P concentrations, which ranged from 0.04-0.30 mg/L. The EPA recommends a minimum of 30 data points be used when analyzing frequency distributions to guide nutrient criteria development (EPA, 2000), but only 2 Texas basins had enough TP and PO_4-P concentrations medians to exceed this threshold for tidal streams. None of the Texas basins had greater than 30 medians for TN concentrations, and frequency 25th percentiles could only be calculated for 7 basins. The 25th percentile of median TN concentrations ranged from 0.75 to 3.5 mg/L for these 7 basins. The 25th percentile of median NO_x-N concentrations was available for 8 basins and ranged from 0.04 to 2.2 mg/L. Maximum median values for each of these nutrient parameters was observed for Basin 22, while the lowest median nutrient concentrations were frequently observed for Basins 5 and 6. The 25th percentile chl-a data could only be calculated for 5 basins due to limited data. The 25th percentile of fluorometric chl-a medians ranged from 4.7 to 16 µg/L. Chl-a was lowest for Basins 6 and 10 and highest for Basin 22.

Table 4.1.1. Frequency distribution by basin of median nutrient and chlorophyll-a concentrations in samples collected from tidal streams in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
5	6	0.06	--	0.073	0.085	0.094	--	0.160
6	4	0.07	--	0.070	0.070	0.073	--	0.080
7	2	0.11	--	--	0.120	--	--	0.130
8	6	0.13	--	0.133	0.163	0.245	--	2.060
9	2	0.295	--	--	0.480	--	--	0.665
10	48	0.11	0.200	0.289	0.543	0.765	0.968	2.310
11	37	0.06	0.092	0.170	0.210	0.305	0.519	1.460

12	1	--	--	--	0.150	--	--	--
13	4	0.22	--	0.303	0.335	0.345	--	0.360
14	1	--	--	--	0.255	--	--	--
15	1	--	--	--	0.220	--	--	--
16	1	--	--	--	0.180	--	--	--
18	1	--	--	--	0.260	--	--	--
20	3	0.1	--	--	0.429	--	--	0.485
21	1	--	--	--	0.190	--	--	--
22	6	0.18	--	0.410	0.455	0.500	--	0.540
23	2	0.368	--	--	0.394	--	--	0.420
24	21	0.1	0.110	0.130	0.200	0.300	0.390	2.790

Total Nitrogen (TN; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
5	1	--	--	--	0.82	--	--	--
6	4	0.72	--	0.75	0.78	0.79	--	0.80
7	2	1.02	--	--	1.16	--	--	1.29
8	5	0.92	--	0.93	1.23	1.81	--	12.40
9	1	--	--	--	1.50	--	--	--
10	22	1.39	1.68	2.45	3.47	5.01	6.22	8.04
11	11	0.91	0.99	1.22	1.55	2.26	2.85	7.51
12	1	--	--	--	1.60	--	--	--
13	4	1.28	--	1.37	1.40	1.72	--	2.67
14	1	--	--	--	1.79	--	--	--
15	1	--	--	--	1.66	--	--	--
16	0	--	--	--	--	--	--	--
18	1	--	--	--	2.69	--	--	--
20	0	--	--	--	--	--	--	--
21	1	--	--	--	1.20	--	--	--
22	6	1.62	--	3.45	3.89	4.36	--	4.41
23	2	1.40	--	--	1.89	--	--	2.37
24	6	0.99	--	1.13	1.32	1.54	--	3.51

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
5	7	0.05	0.06	0.07	0.08	0.10	0.13	0.16
6	4	0.06	--	0.09	0.10	0.10	--	0.11
7	2	0.08	--	--	0.09	--	--	0.10
8	6	0.10	--	0.10	0.21	0.33	--	9.64
9	2	0.04	--	--	0.13	--	--	0.23
10	34	0.05	0.22	0.68	1.63	2.47	3.40	4.66
11	39	0.04	0.04	0.04	0.09	0.25	1.04	5.98
12	1	--	--	--	0.70	--	--	--
13	4	0.12	--	0.17	0.27	0.38	--	0.48
14	1	--	--	--	0.76	--	--	--
15	1	--	--	--	0.28	--	--	--
16	1	--	--	--	0.33	--	--	--
18	1	--	--	--	2.00	--	--	--
20	3	0.02	--	--	0.02	--	--	0.61
21	1	--	--	--	0.10	--	--	--
22	6	0.04	--	2.17	2.66	3.09	--	3.48
23	2	0.13	--	--	0.43	--	--	0.73
24	20	0.02	0.04	0.04	0.06	0.14	0.91	11.81

Ortho-Phosphate (PO_4-P; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
5	7	0.040	0.040	0.040	0.040	0.050	0.086	0.140
6	4	0.060	--	0.060	0.060	0.060	--	0.060
7	2	0.045	--	--	0.053	--	--	0.060
8	6	0.040	--	0.045	0.080	0.130	--	1.800
9	2	0.180	--	--	0.308	--	--	0.435
10	47	0.060	0.106	0.215	0.390	0.720	0.887	1.080
11	41	0.040	0.040	0.060	0.090	0.225	0.380	1.290
12	1	--	--	--	0.060	--	--	--
13	4	0.135	--	0.184	0.205	0.228	--	0.280
14	1	--	--	--	0.141	--	--	--
15	1	--	--	--	0.120	--	--	--
16	1	--	--	--	0.130	--	--	--
18	1	--	--	--	0.170	--	--	--
20	3	0.019	--	--	0.301	--	--	0.430
21	1	--	--	--	0.100	--	--	--
22	6	0.055	--	0.296	0.370	0.433	--	0.475
23	1	--	--	--	0.270	--	--	--
24	19	0.040	0.040	0.040	0.060	0.213	0.620	2.405

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
5	0	--	--	--	--	--	--	--
6	4	5.50	--	5.83	6.39	7.15	--	8.11
7	2	6.82	--	--	10.3	--	--	13.8
8	3	3.00	--	--	11.7	--	--	23.5
9	1	--	--	--	26.2	--	--	--
10	19	3.12	4.18	4.66	5.43	8.08	9.20	10.4
11	11	5.96	6.58	9.78	14.8	18.0	32.9	59.0
12	1	--	--	--	19.8	--	--	--
13	2	10.7	--	--	12.6	--	--	14.5
14	1	--	--	--	7.95	--	--	--
15	1	--	--	--	22.2	--	--	--
16	0	--	--	--	--	--	--	--
18	1	--	--	--	7.69	--	--	--
20	2	5.70	--	--	8.25	--	--	10.8
21	1	25.6	--	--	25.6	--	--	25.6
22	6	14.5	--	15.8	17.6	25.0	--	30.8
23	0	--	--	--	--	--	--	--
24	6	9.97	--	13.0	17.7	18.5	--	20.8

Level III Ecoregion. Texas is divided into 11 level III ecoregions comprised of deserts (9%), tablelands (9%), timbers (9%), plateaus (9%), prairies (9%), and plains (55%) (Appendix 1.8). Texas tidal streams are located primarily within just one of these ecoregions, namely, Level III Ecoregion 34-Western Gulf Coastal Plains. A single tidal streams station was located within Level III Ecoregion 35, insufficient data to calculate 25th percentiles for this ecoregion. For Level III Ecoregion 34, however, ~ 60-150 medians were available for calculating frequency distributions for the 5 parameters of interest. For Level III Ecoregion 34, the 25th percentile of TP and PO_4-P medians was 0.17 mg/L and 0.060 mg/L, respectively (Table 4.1.2). The 25th percentile of TN and NO_X-N medians was 1.29 mg/L and 0.070 mg/L, respectively. These 25th percentiles for median nutrient concentrations were fell in the middle of the range of 25th percentiles estimated for tidal streams by basin. The 25th percentile of fluorometric chl-a medians for Texas tidal streams was 5.96. Level III Ecoregion 34 is located within the Aggregate Nutrient Ecoregion X. The 25th percentile of TP medians for Texas tidal streams in this ecoregion is in range with USEPA recommendations (0.17 mg/L vs. 0.13 mg/L, respectively). The 25th percentile of TN medians, however was ~30% greater than USEPA recommendations (1.29 mg/L vs. 0.76 mg/L), while the 25th percentile of chl-a medians was almost 3 times greater than USEPA recommendations.

The development of frequency distributions from median parameter concentrations is a first step in developing nutrient criteria, and is good method to estimate the number of sites within a spatial scale that could exceed a criteria. However, this study and others (Ice et al. 2003; Binkley 2004; Longing and Haggard 2010) have shown that 25th percentiles can vary between basins or ecoregions and across spatial scales. Regional 25th percentiles often significantly differ from criteria developed for aggregate ecoregions, and USEPA=suggested criteria are often more conservative than regional criteria (Evans-White, 2013). The frequency distribution method should only be one of many tools used to support the development of numeric nutrient criteria. The Science Advisory Board (SAB) has advised the EPA that the stressor-response approach is a legitimate, scientifically based method for developing nutrient criteria when correctly applied, and this approach is the focus of the following sections.

Table 4.1.2. Frequency distribution by level III ecoregion of median nutrient and chlorophyll-a concentrations in samples collected from tidal streams in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	146	0.060	0.100	0.166	0.260	0.489	0.813	2.790
35-South Central Plains	1	--	--	--	0.080	--	--	--

Total Nitrogen (TN; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	68	0.77	0.97	1.29	1.82	3.51	5.15	12.40
35-South Central Plains	1	--	--	--	0.72	--	--	--

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	134	0.02	0.04	0.07	0.18	1.22	2.60	11.81
35-South Central Plains	1	--	--	--	0.06	--	--	--

Phosphate (PO_4-P; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	147	0.019	0.040	0.060	0.165	0.385	0.744	2.405
35-South Central Plains	1	--	--	--	0.060	--	--	--

Fluorometric Chlorophyll-a (Chl-a; μg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
34-Western Gulf Coastal Plain	60	3.00	4.78	5.96	10.3	17.4	23.7	59.0
35-South Central Plains	1	--	--	--	8.11	--	--	--

4.2. GROUPING TIDAL STREAM STATIONS WITH SIMILAR NUTRIENT AND BIOLOGICAL CONDITIONS BY THRESHOLDS IN GEOSPATIAL VARIABLES

Methods

We conducted Categorical and Regression Tree (CART) analyses on the median database for tidal streams (described in Chapter 15) to group reservoir stations by watershed attributes and by similar nutrient and/or biological conditions. The focus (dependent) nutrient variables of these analyses were median TP, TN, and spectrophotometric and fluorometric chlorophyll-a concentrations (TCEQ parameter codes 00665, 00600C, 32211, and 70953, respectively. The watershed attributes considered as predictor (independent) variables were divided into 3 categories: 1) land use/land cover (LULC), 2) waste water treatment plant flow, and 3) regions. LULC was further divided into categories, including percent developed, percent agriculture, percent developed + agriculture, percent forested, and percent wetland. Each of these LULC categories combined several TCEQ LULC codes, as described as part of the development of the geospatial database in Section 4.1. Waste water treatment plant discharge was not weighted by watershed area due to limited availability of watershed area data. Texas tidal streams are located within a single level III ecoregion, therefore, basin was the only region considered in the analysis.

CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is very useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were cross-validated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). For the models with the greatest explanatory power, CART analyses were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model had a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. We did require that all calculated medians have a minimum of ten observations used in calculating the median value (see Section 4.1). A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 4.2, the statistical code and raw output generated for each CART and nCPA analysis conducted for this study on geospatial variability in Texas tidal streams has been compiled.

Results and Discussion

Land use/Land cover

For both nutrients and Secchi transparency, LULC categories were good predictors (r^2=0.11-0.32) in Texas tidal streams. On average, TP and TN concentrations were approximately 50% lower when watershed wetland cover exceeded 8-9.0% (Figs. 4.2.1A-B). Wetlands are known hotspots for physical and biogeochemical removal processes (Mitch and Gosselink 2007), such as denitrification, the transformation of reactive NO_x-N to inert N_2 gas, which removes N from the system via diffusion into the atmosphere. Adsorption sites on soil particles can also intercept P compounds in wetlands, reducing P loading. These findings are congruent with ecological theory, and the identified threshold in wetland cover was in range with recommended wetland to watershed area ratios recommended for constructed wetlands. For TN concentration, the percent developed+agricultural LULC was also a strong predictor (r^2=0.32). When developed+agricultural cover exceeded 85%, TN concentrations were almost 3x higher (Fig. 4.2.2A). For Secchi transparency, percent agricultural land was the strongest LULC predictor variable (Fig. 4.2.2B), with a threshold of 29% agricultural LULC. As the percentage of agricultural land increased in the watershed, Secchi transparency decreased. No statistically significant LULC thresholds were identified for chl-a.

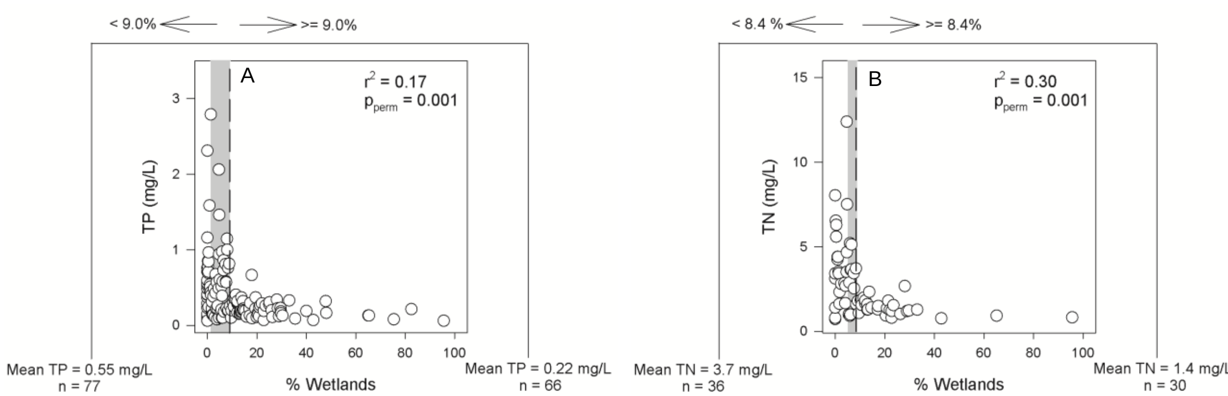


Figure 4.2.1. The relationship between median total phosphorus (TP; A) and total nitrogen (TN; B) and percent wetland land use (% Wetlands) across Texas tidal streams showing thresholds based on classification and regression tree analysis (CART).

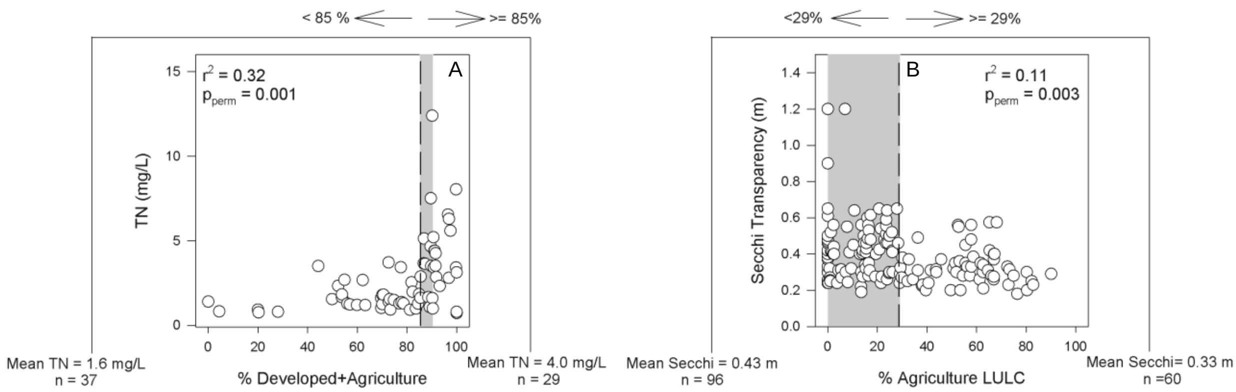


Figure 4.2.2. The relationship between (A) median total nitrogen (TN) and percent developed plus agriculture land use (% Developed+Agriculture) and (B) median Secchi transparency vs. percent agriculture land use (%Agriculture) across Texas tidal streams showing thresholds based on classification and regression tree analysis (CART).

Wastewater Treatment Plant Flow

For Texas tidal streams, permitted municipal WWTP flow was a strong predictor for TP concentration (Fig. 4.2.3.). Above a threshold of 4.6 mgd, TP concentrations more than doubled. This threshold was in range with the unweighted municipal WWTP flow threshold identified in the larger general Texas streams and rivers population and explained a similar amount of variability in tidal streams TP concentrations. However, no other statistically significant thresholds in permitted municipal WWTP flow were identified for TN or chlorophyll-a concentration or Secchi transparency for Texas tidal streams.

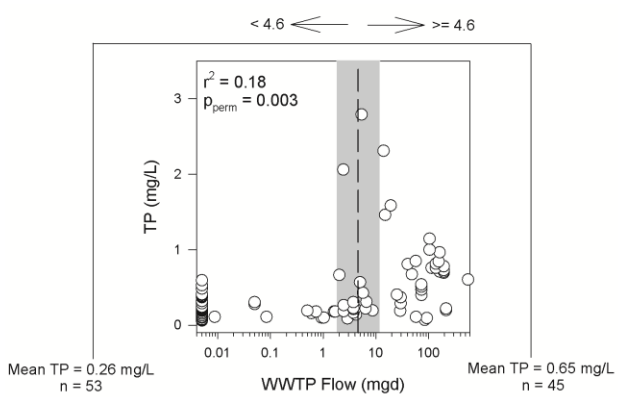


Figure 4.2.3. The relationship between median total phosphorus (TP) and waste water treatment plant (WWTP) flow across Texas tidal streams showing thresholds based on classification and regression tree analysis (CART).

Regions

For Texas tidal streams, CART identified low and high groupings by basin for all variables in the analyses. In general, basin was a strong predictor for these variables (r^2=0.11-0.36), particularly for TN and chl-a. Model explanatory power was lowest for TP compared to other variables and both %Wetland LULC and unweighted permitted municipal WWTP flow explained a larger proportion of data variability. This finding contrasts to findings from the general Texas streams and rivers population and from all other water body types undergoing geospatial analysis, for which regional predictor variables were almost always a very good fit for TP concentrations. The CART groupings are summarized in Table 4.2.1 for TP and TN concentration and 4.2.2. for chlorophyll-a measured using both methods and Secchi transparency.

Table 4.2.1. Groupings of Texas basins by tidal stream stations with "low" and "high" median total phosphorus (TP) and total nitrogen (TN) based on classification and regression tree analysis (CART).

Median TP r^2 = 0.11		Median TN r^2 = 0.24	
"Low", n = 85 Mean TP = 0.28 mg/L	"High", n = 62 Mean TP = 0.55 mg/L	"Low", n = 35 Mean TN = 1.7 mg/L	"High", n = 34 Mean TN = 3.7 mg/L
5	8	5	8
6	9	6	10
7	10	7	18
12	22	9	22
13		11	
14		12	
15		13	
16		14	
18		15	
20		21	
21		23	
23		24	
24			

Table 4.2.2. Groupings of Texas basins by tidal stream stations with "low" and "high" median chl-a and Secchi transparency based on classification and regression tree analysis (CART).

Median Chl-a Spec r^2 = 0.22		Median Chl-a Fluoro r^2 = 0.36		Median Secchi r^2 = 0.14	
"Low", n= 45 Mean= 10 µg/L	"High", n= 28 Mean = 15 µg/L	"Low" n= 34 Mean = 7.7 µg/L	"High" n= 27 Mean = 19 µg/L	"Low" n= 90 Mean = 0.34 m	"High" n= 69 Mean = 0.46 m
6	8	6	9	7	5
7	11	7	11	8	6
9	15	8	12	9	10
10	20	10	15	11	16
12	21	13	21	12	22
13	22	14	22	13	
14	23	18	24	14	
18	8	20		15	
24				18	
				20	
				21	
				23	
				24	

Summary of geospatial analysis in Texas tidal streams

Several geospatial grouping schemes were highly effective for nutrient and biological response variables in the Texas tidal streams water quality database (Table 4.2.3). For total nutrient concentrations, the strongest geospatial models were identified thresholds in %Wetlands and %Developed+Agriculture LULC across all geospatial categories (r^2=0.17-0.32). For TP, models based on %Wetlands and WWTP flow performed similarly. For biological response variables, the most successful geospatial grouping schemes were by basin (r^2=0.14-0.36). In contrast to the general streams and rivers population and other water body types, models identifying geospatial thresholds in TN concentrations had greater explanatory power than for TP. This finding may signal a shift to increased TN limitation in coastal regions.

Table 4.2.3. Summary of classification and regression tree (CART) models with the greatest statistical power in each geospatial category including land use/ land cover (LULC), waste water treatment plant (WWTP) flow, and region across tidal streams in Texas. NS=not significant; Groups=no numerical thresholds.

Parameter	Geospatial Category	Predictor	Threshold	Model r^2
Total Phosphorus	LULC	% Wetlands	9.0%	0.17
	WWTP Flow	Unweighted	4.6 mgd	0.18
	Region	Basin	Groups	0.11
Total Nitrogen	LULC	% Developed+Agriculture	85%	0.32
	WWTP Flow	NS	NS	NS
	Region	Basin	Groups	0.22
Chl-a Spectrophotometric	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Region	Basin	Groups	0.22
Chl-a Fluorometric	LULC	NS	NS	NS
	WWTP Flow	NS	NS	NS
	Region	Basin	Groups	0.36
Secchi Transparency	LULC	% Agriculture	29%	0.11
	WWTP Flow	NS	NS	NS
	Region	Basin	Groups	0.14

4.3. STRESSOR-RESPONSE ANALYSIS ON THE TIDAL STREAMS MEDIAN WATER QUALITY DATABASE

Methods

We conducted CART analyses on the median database for tidal streams (see Section 4.1) to identify thresholds in nutrient concentrations that resulted in measurable changes in common biological responses. The biological (dependent) variables included in the analyses were: median Secchi depth (m; parameter code 00078C), median chlorophyll-a measured with spectrophotometry (chl-a spec; parameter code 32211), and median chlorophyll-a measured with fluorometry (chl-a fluoro; parameter code 70953). The nutrient (independent) variables included in the analysis were total phosphorus (TP; 00665) and total nitrogen (TN; 00600C. Median 24-hour DO Flux was not considered as a response variable in changepoint analyses on the tidal streams median database because there were too few paired observations between DO Flux and nutrients stressors to meet the minimum requirements for CART analysis.

CART analysis is a means to reduce data, based on quantifying thresholds in independent variables that are correlated with shifts in the magnitude and/or variability of dependent variables. This statistical procedure can also provide hierarchical structure in independent variables, showing multiple thresholds from the same or different independent variables. CART analysis is useful for resolving nonlinear, hierarchical, and high-order interactions among predictor variables (De'Ath and Fabricius 2000) and for detecting numerical values that lead to ecological changes (Qian and others 2003). CART models use recursive partitioning to separate data into subsets that are increasingly homogeneous; for example, subsets of data representing similar nutrient conditions. This iterative process invokes a tree-like classification that can reveal relationships that are often difficult to reconcile with conventional linear models (Urban 2002). We "pruned" CART models to generate final models that balanced accuracy within the available dataset with robustness to novel data (Urban 2002). CART models were crossvalidated to determine "pruning size" (i.e., the number of predictor variables included in the model). Model cross-validations were conducted using 10 random and similarly sized subsets of our data according to the method detailed by De'ath and Fabricius (2000). The optimum tree size for each model was selected using the minimum cross-validated error rule (De'ath and Fabricius 2000).

CART analyses were performed using the MVPART library in R 2.9.1 (http://www.r-project.org/). For the models with the greatest explanatory power, CART analyses were followed by non-parametric changepoint analysis in R.2.9.1 to determine model statistical significance and 95% confidence interval about the threshold estimate (Qian et al. 2003, King and Richardson 2003). Non-parametric changepoint analysis uses random permutations to estimate a p value that can be used to determine Type I and II error associated with the threshold. The analysis simultaneously uses bootstrapping to calculate cumulative probability to estimate uncertainty and provide confidence estimates for the threshold. We required a minimum of 20 observations to be used in any single split in the CART model and that each terminal node in the model has a minimum of ten observations. CART analysis is insensitive to missing data. Therefore, we did not remove observations from the data set due to missing values. However, we did require that all calculated medians have a minimum of ten observations used in calculating the median value (see Section 4.1 for details). A user's guide to interpreting CART and nCPA models and associated summary statistics is available in Appendix 1.9. In Appendix 4.3, the statistical code and raw output generated for each CART and nCPA analysis conducted for this study stressor-response relationships in tidal streams has been compiled.

Results and Discussion

Secchi Transparency

For Secchi transparency, a TP threshold of 0.088 mg/L was identified in CART analyses on the Texas tidal streams water quality database (Fig. 4.3.1.A). Only 11 stations had median TP concentration below this threshold. Therefore, the analysis may have been constrained by the range of TP concentrations that were observed in Texas tidal streams. CART identified a threshold in TN concentrations for Secchi transparency, but the model indicated an increase in Secchi transparency with TN concentrations. This finding is not consistent with established ecological theory. A scatterplot of Secchi transparency vs. TN is show in Fig. 4.3.1B. As with TP concentrations, very few median TN concentrations calculated for tidal streams were below the thresholds identified for Secchi transparency in other water body types (~0.5-1.0 mg/L), which may have constrained the analysis.

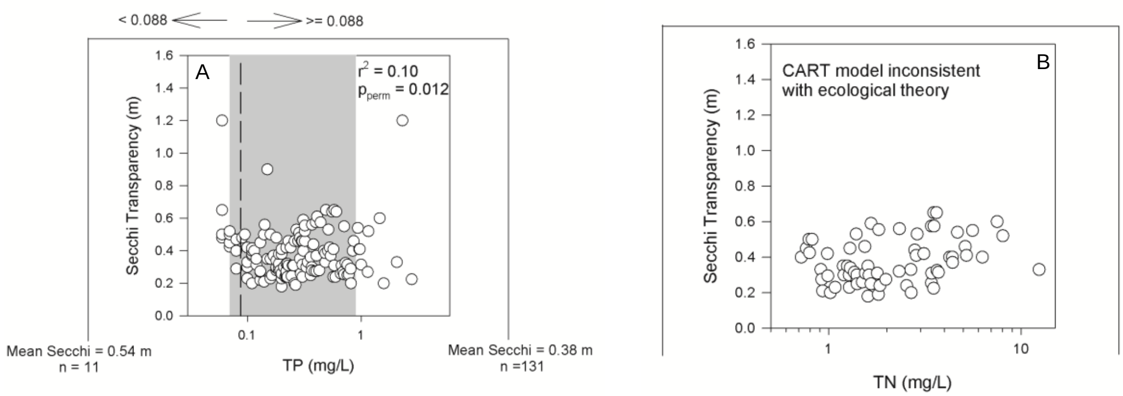


Figure 4.3.1. The relationship between Secchi transparency and (A) total phosphorus (TP) and (B) total nitrogen (TN) across Texas tidal streams showing thresholds based on classification and regression tree analysis (CART).

Chlorophyll-a

No statistically significant thresholds in TP or TN concentrations were observed for chlorophyll-a measured spectrophotometrically or fluorometrically. Scatterplots of chl-a vs. TP and TN are show in Figs. 4.3.2.A-D. For both TP and TN concentrations, only few station medians were less than the nutrient thresholds identified for other water body types, which may have constrained the analysis.

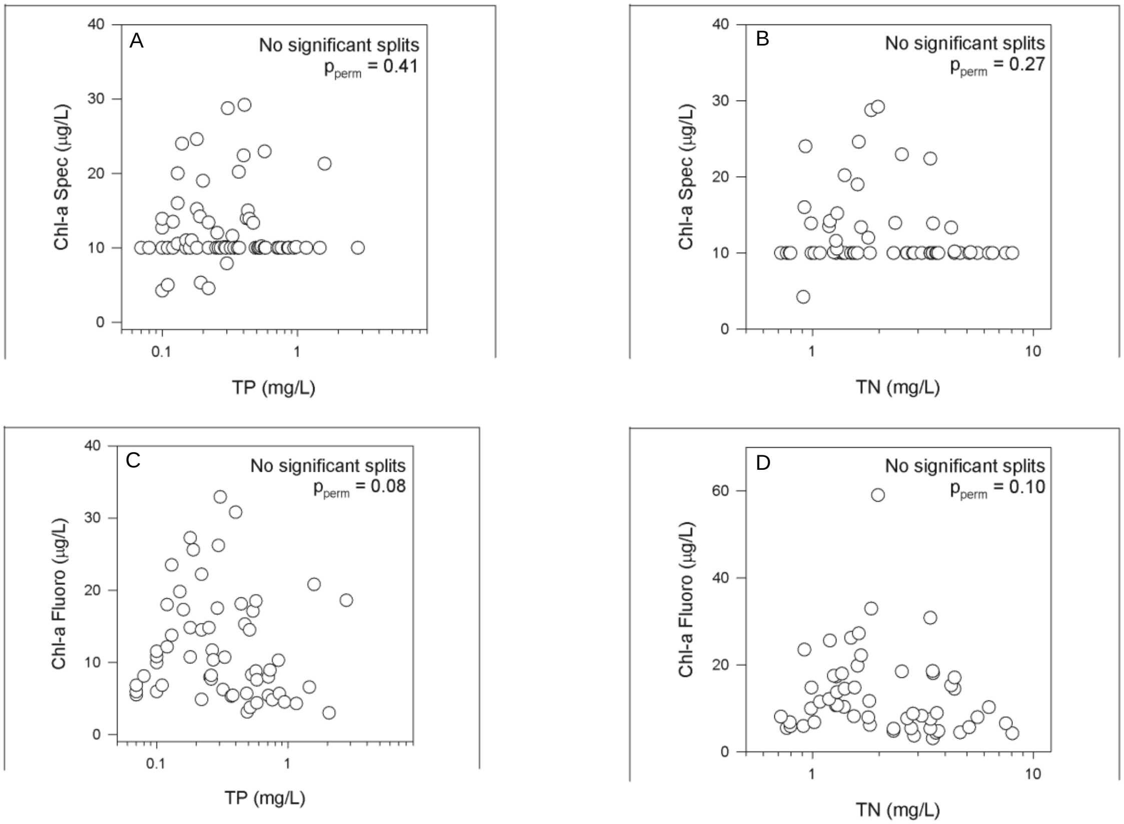


Figure 4.3.2. The relationship between chlorophyll-a measured spectrophotometrically (Chl-a Spec) and (A) total phosphorus (TP) and (B) total nitrogen (TN), as well as chlorophyll-a measured fluorometrically (chl-a fluoro) and (C) TP and (D) TN for the all medians across Texas tidal streams showing thresholds based on classification and regression tree analysis (CART).

Summary of stressor-response analysis in Texas tidal streams

For Texas tidal streams, only one statistically significant and ecologically supported stressor-response model was identified in CART. A TP threshold for Secchi transparency of 0.088 mg/L was in range with TP thresholds identified in the larger streams and rivers median water quality database (~0.06-0.11 mg/L). Data for Texas tidal streams were much more limited than for inland streams and rivers, reservoirs, or estuaries. For inland streams and rivers, relationships between nutrients and chl-a were relatively weak ($r^2 < 0.10$), but trends emerged due to the large volume of data. It is likely that these same patterns are applicable to tidal streams, but the limited availability of data, especially from low nutrient sites, most likely constrained the analyses for this water body type.

REFERENCES

Binkley, D., G.G. Ice, J. Kaye, and C.A. Williams. 2004. Nitrogen and phosphorus concentrations in forest streams of the United States. Journal of American Water Resources Association, 40:1277-1291.Brett et al. 2005

De'ath G, Fabricius KE. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.Dodds et al. 2002

East T L, Sharfstein B. 2006. Development of a decision tree model for the prediction of the limitation potential of phytoplankton in Lake Okeechobee, Florida, USA. Arch. Hydrobiol. 2006;165:127-144.Environmental Protection Agency (EPA). 2000. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002.

Evans-White, M.A., B.E. Haggard, and J.T. Scott. *In Press*. A review of stream nutrient criteria development in the United States. Jounal of Environmental Quality.Gorman, O.T. and J.R. Karr. 1978. Habitat structure and stream fish communities. Ecology. 507-515.

Flum, T., and S.C. Nodvin. 1995. Factors affecting streamwater chemistry in the Great Smoky Mountains, USA. Water, Air, and Soil Pollution, 85(3):1707-1712.

Herlihy AT, Sifneos JC. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. Journal of the North American Benthological Society 27:932-948.

Ice G, Binkley D. 2003. Forest stormwater concentrations of nitrogen and phosphorus: a comparison with EPA's proposed water quality criteria. Journal of Forestry 101:21-28.

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(3):998-1008.

King RS, and CJ Richardson. 2003. Integrating bioasessment and ecological risk assessment: An approach to developing numerical water-quality critera. Environmental Management. 31:795-809. doi:10.1007/s00267-002-0036-4Longing SD, Haggard BE. 2010. Distributions of median nutrient and chlorophyll concentrations across the Red River Basin, USA. Journal of Environmental Quality 39:1966-1974.

Longing, D, and B.E. Haggard. 2010. Distributions of median nutrient and chlorophyll concentrations across the Red River Basin, USA. Journal of Environmental Quality, 39(6): 1966-74.

Mitch, W.J. and J.G. Gosselink. 2007. Wetlands. John Wiley and Sons, Hoboken, NJ.

Qian SS, King RS, Richardson CJ. 2003. Two methods for the detection of environmental thresholds. Ecological Modeling 166:87-97.Robertson et al. 2006

Rohm CM, Omernik JM, Woods AJ, Stoddard JL. 2002. Regional characteristics of nutrient concentrations in streams and their application to nutrient criteria development. Journal of the American Water Resources Association 38:213-239.

Smith RA, Alexander RB, Schwarz GE. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous Unites States. Environmental Science and Technology 37:3039-3047.

Suplee MW, Varghese A, Cleland J. 2007. Developing nutrient criteria for streams: an evaluation of the frequency distribution method. Journal of the American Water Resources Association 43(2):453-472.

APPENDICES

Appendix 4.1. Frequency distributions for Basin x level III and level IV Ecoregions across tidal streams in Texas.

Table A4.1.1: Frequency distribution by basin by level III ecoregion of median nutrient and chlorophyll-a concentrations in samples collected from tidal streams in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
5-34-Western Gulf Coastal Plain	6	0.060	--	0.073	0.085	0.094	--	0.160
6-34-Western Gulf Coastal Plain	3	0.070	--	--	0.070	--	--	0.070
6-35-South Central Plains	1	--	--	--	0.080	--	--	--
7-34-Western Gulf Coastal Plain	2	0.110	--	--	0.120	--	--	0.130
8-34-Western Gulf Coastal Plain	6	0.130	--	0.133	0.163	0.245	--	2.060
9-34-Western Gulf Coastal Plain	2	0.295	--	--	0.480	--	--	0.665
10-34-Western Gulf Coastal Plain	48	0.110	0.200	0.289	0.543	0.765	0.968	2.310
11-34-Western Gulf Coastal Plain	37	0.060	0.092	0.170	0.210	0.305	0.519	1.460
12-34-Western Gulf Coastal Plain	1	--	--	--	0.150	--	--	--
13-34-Western Gulf Coastal Plain	4	0.220	--	0.303	0.335	0.345	--	0.360
14-34-Western Gulf Coastal Plain	1	--	--	--	0.255	--	--	--
15-34-Western Gulf Coastal Plain	1	--	--	--	0.220	--	--	--
16-34-Western Gulf Coastal Plain	1	--	--	--	0.180	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	0.260	--	--	--
20-34-Western Gulf Coastal Plain	3	0.100	--	--	0.429	--	--	0.485
21-34-Western Gulf Coastal Plain	1	--	--	--	0.190	--	--	--
22-34-Western Gulf Coastal Plain	6	0.180	--	0.410	0.455	0.500	--	0.540
23-34-Western Gulf Coastal Plain	2	0.368	--	--	0.394	--	--	0.420
24-34-Western Gulf Coastal Plain	21	0.100	0.110	0.130	0.200	0.300	0.390	2.790

Total Nitrogen (TN; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
5-34-Western Gulf Coastal Plain	1	--	--	--	0.82	--	--	--
6-34-Western Gulf Coastal Plain	3	0.77	--	--	0.79	--	--	0.80
6-35-South Central Plains	1	--	--	--	0.72	--	--	--
7-34-Western Gulf Coastal Plain	2	1.02	--	--	1.16	--	--	1.29
8-34-Western Gulf Coastal Plain	5	0.92	--	0.93	1.23	1.81	--	12.40
9-34-Western Gulf Coastal Plain	1	--	--	--	1.50	--	--	--
10-34-Western Gulf Coastal Plain	22	1.39	1.68	2.45	3.47	5.01	6.22	8.04
11-34-Western Gulf Coastal Plain	11	0.91	0.99	1.22	1.55	2.26	2.85	7.51
12-34-Western Gulf Coastal Plain	1	--	--	--	1.60	--	--	--
13-34-Western Gulf Coastal Plain	4	1.28	--	1.37	1.40	1.72	--	2.67
14-34-Western Gulf Coastal Plain	1	--	--	--	1.79	--	--	--
15-34-Western Gulf Coastal Plain	1	--	--	--	1.66	--	--	--
16-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	2.69	--	--	--
20-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--

21-34-Western Gulf Coastal Plain	1	--	--	--	1.20	--	--	--
22-34-Western Gulf Coastal Plain	6	1.62	--	3.45	3.89	4.36	--	4.41
23-34-Western Gulf Coastal Plain	2	1.40	--	--	1.89	--	--	2.37
24-34-Western Gulf Coastal Plain	6	0.99	--	1.13	1.32	1.54	--	3.51

Nitrite Plus Nitrate-Nitrogen (NO_X-N; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
5-34-Western Gulf Coastal Plain	7	0.05	0.06	0.07	0.08	0.10	0.13	0.16
6-34-Western Gulf Coastal Plain	3	0.10	--	--	0.10	--	--	0.11
6-35-South Central Plains	1	--	--	--	0.06	--	--	--
7-34-Western Gulf Coastal Plain	2	0.11	--	--	0.12	--	--	0.13
8-34-Western Gulf Coastal Plain	6	0.10	--	0.10	0.21	0.33	--	9.64
9-34-Western Gulf Coastal Plain	2	0.04	--	--	0.13	--	--	0.23
10-34-Western Gulf Coastal Plain	34	0.05	0.22	0.68	1.63	2.47	3.40	4.66
11-34-Western Gulf Coastal Plain	39	0.04	0.04	0.04	0.09	0.25	1.04	5.98
12-34-Western Gulf Coastal Plain	1	--	--	--	0.70	--	--	--
13-34-Western Gulf Coastal Plain	4	0.12	--	0.17	0.27	0.38	--	0.48
14-34-Western Gulf Coastal Plain	1	--	--	--	0.76	--	--	--
15-34-Western Gulf Coastal Plain	1	--	--	--	0.28	--	--	--
16-34-Western Gulf Coastal Plain	1	--	--	--	0.33	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	2.00	--	--	--
20-34-Western Gulf Coastal Plain	3	0.02	--	--	0.02	--	--	0.61
21-34-Western Gulf Coastal Plain	1	--	--	--	0.10	--	--	--
22-34-Western Gulf Coastal Plain	6	0.04	--	2.17	2.66	3.09	--	3.48
23-34-Western Gulf Coastal Plain	2	0.13	--	--	0.43	--	--	0.73
24-34-Western Gulf Coastal Plain	20	0.02	0.04	0.04	0.06	0.14	0.91	11.81

Ortho-Phosphate (PO_4-P; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
5-34-Western Gulf Coastal Plain	7	0.040	0.040	0.040	0.040	0.050	0.086	0.140
6-34-Western Gulf Coastal Plain	3	0.060	--	--	0.060	--	--	0.060
6-35-South Central Plains	1	--	--	--	0.060	--	--	--
7-34-Western Gulf Coastal Plain	2	0.045	--	--	0.053	--	--	0.060
8-34-Western Gulf Coastal Plain	6	0.040	--	0.045	0.080	0.130	--	1.800
9-34-Western Gulf Coastal Plain	2	0.180	--	--	0.308	--	--	0.435
10-34-Western Gulf Coastal Plain	47	0.060	0.106	0.215	0.390	0.720	0.887	1.080
11-34-Western Gulf Coastal Plain	41	0.040	0.040	0.060	0.090	0.225	0.380	1.290
12-34-Western Gulf Coastal Plain	1	--	--	--	0.060	--	--	--
13-34-Western Gulf Coastal Plain	4	0.135	--	0.184	0.205	0.228	--	0.280
14-34-Western Gulf Coastal Plain	1	--	--	--	0.141	--	--	--
15-34-Western Gulf Coastal Plain	1	--	--	--	0.120	--	--	--
16-34-Western Gulf Coastal Plain	1	--	--	--	0.130	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	0.170	--	--	--
20-34-Western Gulf Coastal Plain	3	0.019	--	--	0.301	--	--	0.430
21-34-Western Gulf Coastal Plain	1	--	--	--	0.100	--	--	--
22-34-Western Gulf Coastal Plain	6	0.055	--	0.296	0.370	0.433	--	0.475
23-34-Western Gulf Coastal Plain	1	--	--	--	0.270	--	--	--
24-34-Western Gulf Coastal Plain	19	0.040	0.040	0.040	0.060	0.213	0.620	2.405

Fluorometric Chlorophyll-a (Chl-a; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
5-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
6-34-Western Gulf Coastal Plain	3	5.50	--	--	5.94	--	--	6.83
6-35-South Central Plains	1	--	--	--	8.11	--	--	--
7-34-Western Gulf Coastal Plain	2	6.82	--	--	10.3	--	--	13.8
8-34-Western Gulf Coastal Plain	3	3.00	--	--	11.7	--	--	23.5
9-34-Western Gulf Coastal Plain	1	--	--	--	26.2	--	--	--
10-34-Western Gulf Coastal Plain	19	3.12	4.18	4.66	5.43	8.08	9.20	10.4
11-34-Western Gulf Coastal Plain	11	5.96	6.58	9.78	14.8	18.0	32.9	59.0
12-34-Western Gulf Coastal Plain	1	--	--	--	19.8	--	--	--
13-34-Western Gulf Coastal Plain	2	10.7	--	--	12.6	--	--	14.5
14-34-Western Gulf Coastal Plain	1	--	--	--	7.95	--	--	--
15-34-Western Gulf Coastal Plain	1	--	--	--	22.2	--	--	--
16-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
18-34-Western Gulf Coastal Plain	1	--	--	--	7.69	--	--	--
20-34-Western Gulf Coastal Plain	2	5.70	--	--	8.25	--	--	10.8
21-34-Western Gulf Coastal Plain	1	--	--	--	25.6	--	--	--
22-34-Western Gulf Coastal Plain	6	14.5	--	15.8	17.6	25.0	--	30.8
23-34-Western Gulf Coastal Plain	0	--	--	--	--	--	--	--
24-34-Western Gulf Coastal Plain	6	9.97	--	13.0	17.7	18.5	--	20.8

Table A4.1.2. Frequency distribution by basin by level IV ecoregion of median nutrient and chlorophyll-a concentrations in samples collected from tidal streams in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	106	0.060	0.110	0.186	0.270	0.578	0.848	2.310
34b-S Subhumid Glf Cstl Prair	5	0.180	--	0.429	0.485	1.585	--	2.790
34c-Floodplains & Low Terrace	8	0.150	0.199	0.246	0.298	0.345	0.363	0.370
34e-Lower Rio Grande Valley	3	0.440	--	--	0.470	--	--	0.510
34f-Lower Rio Grnd Allv Fldpl	2	0.420	--	--	0.480	--	--	0.540
34g-Texas-Louisiana Cstl Marsh	13	0.060	0.070	0.070	0.095	0.130	0.130	0.265
34h-Mid-Coast Barr Isl C Marsh	7	0.100	0.130	0.165	0.190	0.265	0.294	0.330
34i-Laguna Madre Br Isl C Mrs	2	0.368	--	--	0.384	--	--	0.400
35b-Floodplains & Low Terrace	1	--	--	--	0.080	--	--	--

Total Nitrogen (TN; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	40	0.91	1.07	1.53	2.44	3.68	6.32	12.40
34b-S Subhumid Glf Cstl Prair	2	1.62	--	--	2.56	--	--	3.51
34c-Floodplains & Low Terrace	6	1.25	--	1.40	1.50	1.74	--	2.67
34e-Lower Rio Grande Valley	3	3.52	--	--	4.26	--	--	4.40
34f-Lower Rio Grnd Allv Fldpl	2	2.37	--	--	3.39	--	--	4.41

34g-Texas-Louisiana Cstl Marsh	9	0.77	0.79	0.80	0.92	1.29	1.45	1.81
34h-Mid-Coast Barr Isl C Marsh	4	1.20	--	1.26	1.33	1.71	--	2.69
34i-Laguna Madre Br Isl C Mrs	2	1.40	--	--	2.41	--	--	3.42
35b-Floodplains & Low Terrace	1	--	--	--	0.72	--	--	--

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	94	0.02	0.04	0.05	0.20	1.39	2.60	9.64
34b-S Subhumid Glf Cstl Prair	5	0.02	--	0.04	0.61	1.94	--	11.81
34c-Floodplains & Low Terrace	8	0.04	0.06	0.14	0.27	0.54	0.72	0.76
34e-Lower Rio Grande Valley	3	2.38	--	--	2.94	--	--	3.14
34f-Lower Rio Grnd Allv Fldpl	2	0.73	--	--	2.11	--	--	3.48
34g-Texas-Louisiana Cstl Marsh	13	0.05	0.07	0.08	0.10	0.11	0.13	0.32
34h-Mid-Coast Barr Isl C Marsh	7	0.02	0.03	0.07	0.12	0.45	1.14	2.00
34i-Laguna Madre Br Isl C Mrs	2	0.13	--	--	1.11	--	--	2.10
35b-Floodplains & Low Terrace	1	--	--	--	0.06	--	--	--

Phosphate (PO_4-P; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	108	0.040	0.040	0.068	0.180	0.449	0.783	1.800
34b-S Subhumid Glf Cstl Prair	5	0.055	--	0.301	0.430	1.740	--	2.405
34c-Floodplains & Low Terrace	8	0.060	0.113	0.139	0.200	0.214	0.242	0.280
34e-Lower Rio Grande Valley	3	0.330	--	--	0.410	--	--	0.440
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	0.475	--	--	--
34g-Texas-Louisiana Cstl Marsh	13	0.040	0.040	0.040	0.060	0.060	0.060	0.140
34h-Mid-Coast Barr Isl C Marsh	7	0.019	0.068	0.100	0.130	0.185	0.204	0.210
34i-Laguna Madre Br Isl C Mrs	2	0.270	--	--	0.278	--	--	0.285
35b-Floodplains & Low Terrace	1	--	--	--	0.060	--	--	--

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
34a-N Humid Gulf Cstal Prair	34	3.00	4.32	5.34	8.08	12.0	21.1	59.0
34b-S Subhumid Glf Cstl Prair	4	5.70	--	15.4	19.7	22.4	--	27.3
34c-Floodplains & Low Terrace	4	7.95	--	12.9	16.0	18.1	--	19.8
34e-Lower Rio Grande Valley	3	14.5	--	--	15.3	--	--	18.1
34f-Lower Rio Grnd Allv Fldpl	1	--	--	--	17.1	--	--	--
34g-Texas-Louisiana Cstl Marsh	8	5.50	5.81	6.60	9.27	14.8	19.7	23.5
34h-Mid-Coast Barr Isl C Marsh	5	7.69	--	10.4	10.7	10.8	--	25.6
34i-Laguna Madre Br Isl C Mrs	1	--	--	--	30.8	--	--	--
35b-Floodplains & Low Terrace	1	--	--	--	8.11	--	--	--

Appendix 4.2. Classification and regression tree (CART) and non-parametric changepoint analysis code from geospatial analyses

ANALYSIS: TP VS. ALL NUMERICAL GEOSPATIAL PARAMETERS (CART)

```
Call:
mvpart(form = TP ~ DEV + AG + DEVAG + FOR + WET + FLOW + WFLOW +
    SALINITY, data = tsgeo, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=147 (107 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.26075230      0 1.0000000 1.012815 0.3129405
2 0.08063295      2 0.4784954 1.080563 0.3174321

Node number 1: 147 observations,    complexity param=0.2607523
  mean=0.394602, MSE=0.1637975
  left son=2 (66 obs) right son=3 (77 obs), 4 observations remain
  Primary splits:
      WET   < 0.0896475  to the right, improve=0.16528970, (4 missing)
      FLOW  < 4.575      to the left,  improve=0.15648200, (49 missing)
      DEVAG < 0.8639508  to the left,  improve=0.10356030, (4 missing)
      DEV   < 0.5729372  to the left,  improve=0.09243458, (4 missing)

Node number 2: 66 observations
  mean=0.2176136, MSE=0.01024667

Node number 3: 77 observations,    complexity param=0.2607523
  mean=0.5522597, MSE=0.2506216
  left son=6 (26 obs) right son=7 (31 obs), 20 observations remain
  Primary splits:
      FLOW  < 1.99675    to the left,  improve=0.32724160, (20 missing)
      FOR   < 9.6567e-05 to the left,  improve=0.05242795, (0 missing)
      AG    < 0.416305   to the right, improve=0.03058965, (0 missing)
      DEV   < 0.4326272  to the left,  improve=0.01543942, (0 missing)

Node number 6: 26 observations
  mean=0.2234615, MSE=0.02499186

Node number 7: 31 observations
  mean=0.8917419, MSE=0.3288791
```

ANALYSIS: TP VS. %WETLAND (nCPA)

```
         cp       r2 mean left mean right pperm       5%      25%       50%      75%       95%
[1,] 0.0896475 0.1661467 0.5522597  0.2176136 0.001 0.01378514 0.0815561 0.08762717 0.0896475
0.09271333
```

ANALYSIS: TP VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (nCPA)

```
      cp       r2 mean left mean right pperm    5%     25%   50%  75%     95%
[1,] 4.575 0.1821245 0.2568396  0.6503111 0.003 1.806 1.85125 4.575 4.76 11.8145
```

ANALYSIS: TN VS. ALL NUMERICAL GEOSPATIAL PARAMETERS (CART)

```
Call:
mvpart(form = TN ~ DEV + AG + DEVAG + FOR + WET + FLOW + WFLOW +
   SALINITY, data = tsgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=69 (185 observations deleted due to missingness)

        CP nsplit rel error    xerror      xstd
1 0.3246441      0  1.000000 1.0337751 0.3572612
2 0.1076008      1  0.675356 0.9245202 0.2916564

Node number 1: 69 observations,    complexity param=0.3246441
  mean=2.644783, MSE=4.218018
  left son=2 (37 obs) right son=3 (29 obs), 3 observations remain
  Primary splits:
      DEVAG < 0.8548462  to the left,  improve=0.3132944, (3 missing)
      WET   < 0.08440367 to the right, improve=0.3001724, (3 missing)
      DEV   < 0.1697896  to the left,  improve=0.1751419, (3 missing)
      FOR   < 0.06792236 to the right, improve=0.1141016, (3 missing)

Node number 2: 37 observations
  mean=1.646149, MSE=0.5584488

Node number 3: 29 observations
  mean=4.014397, MSE=6.065351
```

ANALYSIS: TN VS. % WETLAND (nCPA)

```
          cp        r2 mean left mean right pperm       5%       25%       50%       75%       95%
[1,] 0.08440367 0.3036184  3.737014   1.426417 0.001 0.0510405 0.05360383 0.0831765 0.08440367
0.08979037
```

ANALYSIS: TN VS. %DEVELOPED+AGRICULTURE

```
         cp       r2 mean left mean right pperm       5%       25%       50%       75%       95%
[1,] 0.8548462 0.316891  1.646149   4.014397 0.001 0.8520173 0.8582006 0.8617016 0.8935299
0.9050595
```

ANALYSIS: TN VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (nCPA)

```
       cp       r2 mean left mean right pperm     5%    25%   50%     75%    95%
[1,] 1.7117 0.1378364     1.557   3.524038 0.111 1.5635 1.7117 4.575 29.79355 65.049
```

ANALYSIS: SECCHI TRANSPARENCY VS. ALL NUMERICAL GEOSPATIAL PARAMETERS

```
Call:
mvpart(form = SECCHI ~ DEV + AG + DEVAG + FOR + WET + FLOW +
   WFLOW + SALINITY, data = tsgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=159 (95 observations deleted due to missingness)

         CP nsplit rel error   xerror      xstd
1 0.1371782      0 1.0000000 1.004899 0.2495191
2 0.1242225      1 0.8628218 1.150903 0.2835636

Node number 1: 159 observations,    complexity param=0.1371782
  mean=0.3912264, MSE=0.0238279
  left son=2 (90 obs) right son=3 (69 obs)
  Primary splits:
      FOR   < 0.003095761 to the right, improve=0.11089500, (4 missing)
      AG    < 0.2866012   to the right, improve=0.10642430, (4 missing)
      WET   < 0.001588335 to the right, improve=0.10486200, (4 missing)
      DEVAG < 0.9966223   to the left,  improve=0.08220439, (4 missing)

Node number 2: 90 observations,    complexity param=0.1242225
  mean=0.3411667, MSE=0.02077558
  left son=4 (78 obs) right son=5 (10 obs), 2 observations remain
```

```
 Primary splits:
    WET   < 0.002512844 to the right, improve=0.24717450, (2 missing)
    DEV   < 0.7023287   to the left,  improve=0.22219380, (2 missing)
    AG    < 0.1821726   to the right, improve=0.16511620, (2 missing)
    FOR   < 0.008641759 to the right, improve=0.15531420, (2 missing)
    DEVAG < 0.9138341   to the left,  improve=0.08035018, (2 missing)

Node number 3: 69 observations
 mean=0.4565217, MSE=0.02027703

Node number 4: 78 observations
 mean=0.3161538, MSE=0.007269822

Node number 5: 10 observations
 mean=0.5445, MSE=0.08321225
```

ANALYSIS: SECCHI TRANSPARENCY VS. %AGRICULTURE (nCPA)

```
        cp        r2 mean left mean right pperm        5%       25%       50%       75%       95%

[1,] 0.2866012 0.1080654 0.4312105     0.3265 0.003 0.001256282 0.2484085 0.2840102 0.2866012
0.2908094
```

ANALYSIS: SECCHI TRANSOARENCY VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (nCPA)

```
        cp         r2 mean left mean right pperm    5%      25%      50%      75%      95%

[1,] 177.8404 0.02739809 0.3883854  0.4642857 0.151 0.067 86.58523 158.2615 214.6084 386.1946
```

ANALYSIS: CHL-A SPEC VS. ALL NUMERICAL GEOSPATIAL PARAMETERS (CART)

```
Call:
mvpart(form = CHLASPEC ~ DEV + AG + DEVAG + FOR + WET + FLOW +
   WFLOW + SALINITY, data = tsgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=73 (181 observations deleted due to missingness)

       CP nsplit rel error   xerror      xstd
1 0.21538637      0 1.0000000 1.020158 0.2565425
2 0.07588281      1 0.7846136 1.170032 0.2966107

Node number 1: 73 observations,    complexity param=0.2153864
 mean=12.2062, MSE=25.00597
 left son=2 (45 obs) right son=3 (28 obs)
```

```
 Primary splits:
   SALINITY < 1.5      to the left,  improve=0.06861127, (25 missing)
    DEV     < 0.645989  to the right, improve=0.06653008, (3 missing)
    FOR     < 0.102364  to the right, improve=0.06489107, (3 missing)
    AG      < 0.6212373 to the left,  improve=0.05731263, (3 missing)

Node number 2: 45 observations
 mean=10.37556, MSE=6.323069

Node number 3: 28 observations
 mean=15.1483, MSE=40.99012
```

ANALYSIS: CHL-A SPEC VS. %FOREST (nCPA)

```
       cp       r2 mean left mean right pperm      5%      25%     50%      75%      95%

[1,] 0.102364 0.06535209    13.125   10.32284 0.207 0.03712814 0.06135121 0.102364 0.1115474
0.1410426
```

ANALYSIS: CHL-A SPEC VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (nCPA)

```
     cp       r2 mean left mean right pperm    5%     25%    50%    75%     95%

[1,] 29.184 0.05556391  13.41533   10.87187 0.503 0.6635 4.079375 27.2975 41.0829 78.93605
```

ANALYSIS: CHL-A FLUORO VS. ALL NUMERICAL GEOSPATIAL PARAMETERS (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ DEV + AG + DEVAG + FOR + WET + FLOW +
   WFLOW + SALINITY, data = tsgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
 n=61 (193 observations deleted due to missingness)

         CP nsplit rel error    xerror      xstd
1 0.36390138      0 1.0000000 1.0327405 0.4168583
2 0.05630814      1 0.6360986 0.8120383 0.3430332

Node number 1: 61 observations,    complexity param=0.3639014
 mean=12.73164, MSE=87.59399
 left son=2 (34 obs) right son=3 (27 obs)
 Primary splits:
    DEV   < 0.6652825  to the right, improve=0.1717298, (1 missing)
    AG    < 0.1191993  to the left,  improve=0.1065560, (1 missing)
```

```
    WET   < 0.09955619 to the left,  improve=0.0974821, (1 missing)
    DEVAG < 0.8548462  to the right, improve=0.0608680, (1 missing)

Node number 2: 34 observations
  mean=7.700441, MSE=15.50474

Node number 3: 27 observations
  mean=19.06722, MSE=106.3579
```

ANALYSIS: CHL-A FLUORO VS. %AGRICULTURE (nCPA)

```
         cp       r2 mean left mean right pperm       5%      25%      50%      75%      95%

[1,] 0.1191993 0.1066711  6.610417   14.31156 0.084 0.1063779 0.1192338 0.1786217 0.5993129
0.6613108
```

ANALYSIS: CHL-A FLUORO VS. PERMITTED MUNICIPAL WASTEWATER TREATMENT PLANT FLOW UNWEIGHTED (nCPA)

```
       cp        r2 mean left mean right pperm     5%   25%   50%    75%      95%
[1,] 17.04 0.04503893  10.81412   15.37529 0.644 0.8752 5.185 20.05 43.475 78.93605
```

ANALYSIS: TP VS. BASIN (CART)

```
Call:
mvpart(form = TP ~ Basin, data = tidalgeo, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=147 (107 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.10558114      0 1.0000000 1.014512 0.3131991
2 0.02324305      1 0.8944189 1.065782 0.3352845

Node number 1: 147 observations,    complexity param=0.1055811
  mean=0.394602, MSE=0.1637975
  left son=2 (85 obs) right son=3 (62 obs)
  Primary splits:
      Basin splits as  RLLLLLLLLLRLLLLLRR, improve=0.1055811, (0 missing)

Node number 2: 85 observations
  mean=0.2822882, MSE=0.1318974

Node number 3: 62 observations
  mean=0.5485806, MSE=0.166528
```

ANALYSIS: TN VS. BASIN (CART)

```
Call:
mvpart(form = TN ~ Basin, data = tidalgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=69 (185 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.23547978      0 1.0000000 1.027965 0.3523594
2 0.01807555      1 0.7645202 1.004365 0.4124288

Node number 1: 69 observations,    complexity param=0.2354798
  mean=2.644783, MSE=4.218018
  left son=2 (35 obs) right son=3 (34 obs)
  Primary splits:
      Basin splits as  RLLLLL-R-LRLLLLLRL, improve=0.2354798, (0 missing)

Node number 2: 35 observations
  mean=1.6625, MSE=1.402888

Node number 3: 34 observations
  mean=3.655956, MSE=5.100216
```

ANALYSIS SECCHI TRANSPARENCY VS. BASIN (CART)

```
Call:
mvpart(form = SECCHI ~ Basin, data = tidalgeo, xval = 10, method = "anova",
   minsplit = 20, minbucket = 10)
  n=159 (95 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.13717817      0 1.0000000 1.0115698 0.2521069
2 0.02926722      1 0.8628218 0.9782254 0.2713579
3 0.01000000      2 0.8335546 0.9533708 0.2696029

Node number 1: 159 observations,    complexity param=0.1371782
  mean=0.3912264, MSE=0.0238279
  left son=2 (90 obs) right son=3 (69 obs)
  Primary splits:
      Basin splits as  RLLLLLRLLLRLLRRLLL, improve=0.1371782, (0 missing)

Node number 2: 90 observations,    complexity param=0.02926722
  mean=0.3411667, MSE=0.02077558
  left son=4 (20 obs) right son=5 (70 obs)
  Primary splits:
```

```
    Basin splits as  -RRLLL-LRL-LR--RLL, improve=0.05930191, (0 missing)

Node number 3: 69 observations
  mean=0.4565217, MSE=0.02027703

Node number 4: 20 observations
  mean=0.2755, MSE=0.00192225

Node number 5: 70 observations
  mean=0.3599286, MSE=0.02457821
```

ANALYSIS: CHL-A SPEC VS. BASIN (CART)

```
Call:
mvpart(form = CHLASPEC ~ DEV + AG + DEVAG + FOR + WET + FLOW +
    WFLOW + SALINITY + Basin, data = tsgeo, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=73 (181 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.21538637      0 1.0000000 1.020158 0.2565425
2 0.07588281      1 0.7846136 1.170032 0.2966107

Node number 1: 73 observations,    complexity param=0.2153864
  mean=12.2062, MSE=25.00597
  left son=2 (45 obs) right son=3 (28 obs)
  Primary splits:
      Basin    splits as  LRLLLR-LRRRRL-LLRL, improve=0.21538640, (0 missing)
      SALINITY < 1.5       to the left,  improve=0.06861127, (25 missing)
      DEV      < 0.645989  to the right, improve=0.06653008, (3 missing)
      FOR      < 0.102364  to the right, improve=0.06489107, (3 missing)
      AG       < 0.6212373 to the left,  improve=0.05731263, (3 missing)

Node number 2: 45 observations
  mean=10.37556, MSE=6.323069

Node number 3: 28 observations
  mean=15.1483, MSE=40.99012
```

ANALYSIS: CHL-SPEC VS. BASIN (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ basin, data = tsgeo, xval = 10, method = "anova",
    minsplit = 20, minbucket = 10)
  n=61 (193 observations deleted due to missingness)
```

```
          CP nsplit rel error    xerror      xstd
1 0.36390138      0 1.0000000 1.0327405 0.4168583
2 0.05630814      1 0.6360986 0.8120383 0.3430332

Node number 1: 61 observations,    complexity param=0.3639014
  mean=12.73164, MSE=87.59399
  left son=2 (34 obs) right son=3 (27 obs)
  Primary splits:
      Basin splits as  LRRLLR-LLRR-R-LLLR, improve=0.3639014, (0 missing)

Node number 2: 34 observations
  mean=7.700441, MSE=15.50474

Node number 3: 27 observations
  mean=19.06722, MSE=106.3579
```

Appendix 4.3. Classification and regression tree (CART) and non-parametric changepoint analysis code from geospatial analyses

ANALYSIS: SECCHI TRANSPARENCY VS. NUTRIENT STRESSORS (TP, TN, NH_4-N, NO_X-N, SRP) (CART)

```
Call:
mvpart(form = SECCHI ~ TP + TN + NH4 + NOX + SRP, data = tidal,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=157 (97 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.24870331      0 1.0000000 1.010044 0.2528330
2 0.07492739      1 0.7512967 1.012320 0.2748852

Node number 1: 157 observations,    complexity param=0.2487033
  mean=0.3921338, MSE=0.02406089
  left son=2 (106 obs) right son=3 (29 obs), 22 observations remain
  Primary splits:
      NH4 < 0.1825   to the left,  improve=0.14833420, (22 missing)
      NOX < 0.5175   to the left,  improve=0.08875532, (34 missing)
      TP  < 0.0975   to the right, improve=0.07474906, (15 missing)
      TN  < 3.5075   to the left,  improve=0.04794503, (91 missing)

Node number 2: 106 observations
  mean=0.3533019, MSE=0.01914193

Node number 3: 29 observations
  mean=0.5101724, MSE=0.02789738
```

ANALYSIS: SECCHI TRANSPARENCY VS. TP (nCPA)

```
         cp         r2 mean left mean right pperm   5%  25%    50%   75% 95%

[1,] 0.0875 0.09829581     0.535   0.375229 0.012 0.07 0.08 0.0975 0.265 0.9
```

ANALYSIS: SECCHI TRANSPARENCY VS. TN (nCPA)

```
         cp        r2 mean left mean right pperm   5%      25%    50%    75%    95%

[1,] 3.5075 0.1796102    0.3462  0.4684375 0.004 1.64 2.731875 2.7675 3.5075 3.5675
```

ANALYSIS: SECCHI TRANSPARENCY VS. NH_4-N (nCPA)

```
         cp        r2 mean left mean right pperm   5%    25%    50%     75%   95%
```

[1,] 0.0725 0.0385738 0.4545455 0.3754167 0.303 0.06 0.0725 0.0875 0.12625 0.255

ANALYSIS: CHL-A SPEC VS. NUTRIENT STRESSORS (TP, TN, NH4-N, NOX-N, SRP) (CART)

```
Call:
mvpart(form = CHLASPEC ~ TP + TN + NH4 + NOX + SRP, data = tidal,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=73 (181 observations deleted due to missingness)

          CP nsplit rel error    xerror      xstd
1 0.26383428      0 1.0000000 1.0238963 0.2571623
2 0.03205123      1 0.7361657 0.8494233 0.2458774

Node number 1: 73 observations,    complexity param=0.2638343
  mean=12.2062, MSE=25.00597
  left son=2 (51 obs) right son=3 (19 obs), 3 observations remain
  Primary splits:
      NH4 < 0.055  to the right, improve=0.12362920, (3 missing)
      TN  < 2.605  to the right, improve=0.05385218, (10 missing)
      NOX < 0.695  to the right, improve=0.04583057, (1 missing)
      SRP < 0.0575 to the right, improve=0.04423229, (3 missing)
      TP  < 0.115  to the left,  improve=0.04362129, (0 missing)

Node number 2: 51 observations
  mean=10.73338, MSE=13.00052

Node number 3: 19 observations
  mean=14.77105, MSE=35.8314
```

ANALYSIS: CHL-A SPEC VS. TP (nCPA)

```
        cp         r2 mean left mean right pperm    5%   25% 50%      75%   95%
[1,] 0.115 0.04362129   9.58475   12.62230 0.413 0.115 0.125 0.3 0.460875 0.575
```

ANALYSIS: CHL-A SPEC VS. TN (nCPA)

```
        cp         r2 mean left mean right pperm    5%    25%   50%   75%   95%
[1,] 2.605 0.06583317  13.35125     10.798 0.268 1.135 2.1525 2.605 3.425 3.575
```

ANALYSIS: CHL-A SPEC VS. NH_4-N (nCPA)

```
        cp        r2 mean left mean right pperm    5%   25%   50%   75%   95%
[1,] 0.055 0.1437891  14.77105   10.73338 0.019 0.055 0.055 0.055 0.105 0.125
```

ANALYSIS: CHL-A FLUORO VS. NUTRIENT STRESSORS (TP, TN, NH4-N, NOX-N, SRP) (CART)

```
Call:
mvpart(form = CHLAFLUORO ~ TP + TN + NH4 + NOX + SAL + SRP, data = tidal,
   xval = 10, method = "anova", minsplit = 20, minbucket = 10)
  n=61 (193 observations deleted due to missingness)

          CP nsplit rel error   xerror      xstd
1 0.17404314      0 1.0000000 1.013121 0.4076764
2 0.05533397      1 0.8259569 1.028099 0.3682577

Node number 1: 61 observations,    complexity param=0.1740431
  mean=12.73164, MSE=87.59399
  left son=2 (45 obs) right son=3 (16 obs)
  Primary splits:
      NH4 < 0.0525 to the right, improve=0.1740431, (0 missing)
      NOX < 0.4425 to the right, improve=0.1251472, (0 missing)
      SRP < 0.2875 to the right, improve=0.1115897, (0 missing)
      TN  < 2.1525 to the right, improve=0.1065138, (3 missing)
      TP  < 0.4775 to the right, improve=0.1019443, (0 missing)

Node number 2: 45 observations
  mean=10.40344, MSE=41.00948

Node number 3: 16 observations
  mean=19.27969, MSE=160.4909
```

ANALYSIS: CHL-A FLUORO VS. TP (nCPA)

```
         cp        r2 mean left mean right pperm    5%   25%    50%  75%   95%

[1,] 0.4775 0.1019443  14.97603   8.752955 0.077 0.115 0.445 0.4775 0.48 0.575
```

ANALYSIS: CHL-A FLUORO VS. TN (nCPA)

```
         cp        r2 mean left mean right pperm    5%    25%    50%  75%   95%

[1,] 2.1525 0.1089286  15.67032    9.39037 0.098 1.195 2.0875 2.1525 2.67 3.575
```

ANALYSIS: CHL-A FLUORO VS. NH4-N (nCPA)

```
         cp        r2 mean left mean right pperm     5%    25%    50%   75%  95%

[1,] 0.0525 0.1740431  19.27969   10.40344 0.006 0.0525 0.0525 0.0875 0.115 0.15
```

Section 5: Springs

E.M. Grantz, L.B. Massey, B.E. Haggard, J.T. Scott

EXECUTIVE SUMMARY

The Clean Water Act directs state to develop numeric nutrient criteria for all water bodies, including streams and rivers, lakes and reservoirs, estuaries and coastal waters, and even wetlands. However, groundwater and springs have not been part of these discussions nor are these water bodies specifically mentioned in the context of nutrient criteria development. Inadvertently, watershed management to meet numeric nutrient criteria in surface waters (i.e., fresh and coastal) will address groundwater and springs, because of the interconnection with surface waters. The chemistry of groundwater and springs play a strong role in the chemical conditions of surface waters, especially under base flow conditions when groundwater inputs make up a large proportion of flow. However, biogeochemical process in the riparian and hyporheic zones can have a profound influence on nutrient dynamics, changing the chemistry from groundwater to that expressed in surface waters.

The goal of this chapter was to examine the frequency distribution of median nutrient concentrations across a limited number of Texas springs, where these data were provided by the Texas Commission on Environmental Quality (TCEQ). These frequency distributions are presented in Section 5.1. Available data were limited regardless of the spatial scale considered. Therefore, no estimates of 25th percentiles could be made, though estimation of minimum, maximum, and median values was possible. Also due to the lack of information available for springs, analyses to evaluate stressor response information could not be performed. Additional data collection efforts would be necessary before the water quality of Texas springs could be reliably evaluated for setting criteria.

INTRODUCTION

The Clean Water Act directs state to develop numeric nutrient criteria for all water bodies, including streams and rivers, lakes and reservoirs, estuaries and coastal waters, and even wetlands. However, groundwater and springs have not been part of these discussions nor are these water bodies specifically mentioned in the context of nutrient criteria development. Inadvertently, watershed management to meet numeric nutrient criteria in surface waters (i.e., fresh and coastal) will address groundwater and springs, because of the interconnection with surface waters. The chemistry of groundwater and springs play a strong role in the chemical conditions of surface waters, especially under base flow conditions when groundwater inputs make up a large proportion of flow. However, biogeochemical process in the riparian and hyporheic zones can have a profound influence on nutrient dynamics, changing the chemistry from groundwater to that expressed in surface waters.

The goal of this chapter is to examine the frequency distribution of median nutrient concentrations across a limited number of Texas springs, where these data were provided by the Texas Commission on Environmental Quality (TCEQ). Due to the lack of information available for springs, analyses to evaluate stressor response information could not be performed.

5.1: SPRINGS DATABASE DEVELOPMENT, MEDIAN CALCULATION, AND FREQUENCY DISTRIBUTIONS

METHODS

Water Quality Database

Data Acquisition, Compilation and Reduction. TCEQ provided a database of water quality data collected from 1968 to 2012 from springs located in Texas. The collected data was from six stations spanning five watersheds. The data described 116 springs characteristics and water quality parameters including nutrients, sediments, transparency, physico-chemical parameters, as well as others.

For the purposes of advanced statistical analyses conducted during this project, only data collected under specific monitoring codes (as decided by TCEQ) and from 2000 to 2010 was used. Therefore, the database was sorted and any data collected before calendar year 2000 or after 2010 was removed. Data collected under the monitoring code Biased Flow (BF) was also removed since data collected under this circumstance were not necessarily representative of baseline water quality conditions. The data received from TCEQ were output to a single column format within the files, so the data were reorganized into a useable format. The data was sorted by Basin ID and a new Microsoft Excel worksheet was created for each individual basin. Each basin worksheet was then restructured using the pivot table function in Microsoft Excel so that each parameter and the associated data were unique to an individual column; a portion of this process was accomplished with a Mircosoft Excel Macro (see Appendix 1.1 for Excel Macro code). Any estimated data points (i.e., those reported with a < or >) were flagged and used in the database without the associated qualifying sign. The data were flagged using a Microsoft Excel Macro (Appendix 1.2).

Several additional parameters were calculated from the original data provided. Nitrate plus nitrite and total N (TN) were calculated if the necessary N species were provided by TCEQ in the original data file. In addition, diel change (i.e., 24 hour maximum minus 24 hour minimum) was calculated for dissolved oxygen, temperature, conductivity, pH, and turbidity. The additional parameters were added to each station worksheet using a Microsoft Excel Macro (Appendix 1.3).

Where applicable and similarly to procedures undertaken for other water body types, several parameters were removed from the median database because of lack of data and duplication of parameters, or because TCEQ indicated that the parameter could be removed from the database.

Median and Frequency Distribution Calculations. For this study, frequency distribution and, subsequently, stressor-response analyses were conducted on station medians in order to focus on broadly applicable regional and statewide trends. Because each spring in Texas was not equally represented in the raw water quality dataset, conducting statistical analyses on medians removes potential site-specific bias for sites that are over- or under-represented in the raw dataset. Furthermore, biological response and nutrient stressor data did not always overlap in the raw data. Median values of each parameter were calculated for each Station ID using a Microsoft Excel Macro (Appendix 1.4). Median values were calculated based on at least 10 data points, i.e. no medians were calculated if less than 10 data points were available for a given parameter at a given station. The calculated medians for each Station ID were then compiled into one database using a Microsoft Excel Macro (Appendix 1.5). This database was merged with the GIS and LULC data and used in advanced statistical analysis.

Frequency distributions (minimum value, 10th, 25th, 50th, 75th, 90th percentiles and maximum value) for water quality parameters TP (TCEQ parameter code 00665), TN (calculated parameter code 00600C; TCEQ parameter code 00625 + 00630, 00625 + 00593 or 00625 + 00615 + 00620), NO_x-N (calculated parameter 00630C; TCEQ parameter code 00630, 00593 or 00615 + 00620), and PO_4-P (TCEQ calculated parameter code 00671C; TCEQ parameter code 00671 or 70507). For Texas springs, no chlorophyll-a data were available. Frequency distributions were calculated for the springs population at multiple spatial scales including basin, level III ecoregion, basin by level III ecoregion (i.e., unique combinations of basin and level III ecoregions combined), and level IV ecoregion.

Data Quality Assurance and Control

Data quality checks were employed frequently throughout the database reorganization and data calculation processes. The original source files were maintained in an unaltered form, and subsequent changes to each database were saved under unique file names. Data transferred from one file to the next were checked for accuracy by comparing first and last rows and the row count between files. In addition, when calculations were preformed, including manual calculations and those calculated using

Microsoft Excel Macros, at least 10 percent of calculations were checked for accuracy following the secondary data quality assurance project plan (QAPP).

Results and Discussion

Available data for analyzing frequency distributions in Texas springs was extremely limited regardless of the spatial scale considered. Therefore, it was not possible to estimate 25th percentiles of nutrient fractions. For basins and level III ecoregion containing one or more observations for a given parameter, minimum, maximum and median nutrient concentrations are reported in Tables 5.1.1. and 5.2.2., respectively. See Appendix 5.1 for analyses conducted at the basin by level III ecoregion and level IV ecoregion scale.

Table 5.1.1. Frequency distribution by basin of median nutrient concentrations in samples collected from springs in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
12	0	--	--	--	--	--	--	--
14	3	0.02	--	--	0.06	--	--	0.06
18	1	--	--	--	0.05	--	--	--
19	2	0.021	--	--	0.022	--	--	0.023
23	0	--	--	--	--	--	--	--

Total Nitrogen (TN; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
12	0	--	--	--	--	--	--	--
14	3	1.55	--	--	7.04	--	--	7.21
18	0	--	--	--	--	--	--	--
19	2	2.04	--	--	2.11	--	--	2.17
23	0	--	--	--	--	--	--	--

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
12	0	--	--	--	--	--	--	--
14	3	1.39	--	--	6.77	--	--	7.01
18	1	--	--	--	0.24	--	--	--
19	2	1.82	--	--	1.88	--	--	1.93
23	0	--	--	--	--	--	--	--

Ortho-Phosphate (PO_4-P; mg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
12	0	--	--	--	--	--	--	--
14	3	0.02	--	--	0.025	--	--	0.04
18	0	--	--	--	--	--	--	--
19	2	0.02	--	--	0.02	--	--	0.02
23	0	--	--	--	--	--	--	--

Fluorometric Chlorophyll-a (Chl-a; µg/L)

Basin	Count	MIN	10th	25th	Median	75th	90th	MAX
NA								

Table 5.1.2. Frequency distribution by basin by level III ecoregion of median nutrient concentrations in samples collected from springs in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
12-29-Cross Timbers	0	--	--	--	--	--	--	--
14-30-Edwards Plateau	2	0.060	--	--	0.060	--	--	0.060
14-32-Texas Blackland Prairies	1	--	--	--	0.020	--	--	--
18-30-Edwards Plateau	1	--	--	--	0.050	--	--	--
19-32-Texas Blackland Prairies	2	0.021	--	--	0.022	--	--	0.023
23-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--

Total Nitrogen (TN; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
12-29-Cross Timbers	0	--	--	--	--	--	--	--
14-30-Edwards Plateau	2	7.04	--	--	7.13	--	--	7.21
14-32-Texas Blackland Prairies	1	--	--	--	1.55	--	--	--
18-30-Edwards Plateau	0	--	--	--	--	--	--	--
19-32-Texas Blackland Prairies	2	2.04	--	--	2.11	--	--	2.17
23-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--

Nitrite Plus Nitrate-Nitrogen (NO_X-N; mg/L)

Basin-Level III	Count	MIN	10th	25th	Median	75th	90th	MAX
12-29-Cross Timbers	0	--	--	--	--	--	--	--
14-30-Edwards Plateau	2	6.77	--	--	6.89	--	--	7.01
14-32-Texas Blackland Prairies	1	--	--	--	1.39	--	--	1.39
18-30-Edwards Plateau	1	--	--	--	0.24	--	--	--
19-32-Texas Blackland Prairies	2	1.82	--	--	1.88	--	--	1.93
23-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--

Ortho-Phosphate (PO_4-P; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
12-29-Cross Timbers	0	--	--	--	--	--	--	--
14-30-Edwards Plateau	2	0.025	--	--	0.033	--	--	0.040
14-32-Texas Blackland Prairies	1	--	--	--	0.020	--	--	--
18-30-Edwards Plateau	0	--	--	--	--	--	--	--
19-32-Texas Blackland Prairies	2	0.020	--	--	0.020	--	--	0.020
23-24-Chihuahuan Deserts	0	--	--	--	--	--	--	--

Chlorophyll-a (Chl-a; mg/L)

Basin-Level III	Count	MIN	10^{th}	25th	Median	75th	90th	MAX
NA								

APPENDICES

Appendix 5.1. Frequency distributions of springs data for basin x level III and level IV Ecoregions.

Table A5.1.1. Frequency distribution by basin by level III ecoregion of median nutrient concentrations in samples collected from springs in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
29-Cross Timbers	0	--	--	--	--	--	--	--
30-Edwards Plateau	3	0.05	--	--	0.06	--	--	0.06
32-Texas Blackland Prairies	3	0.02	--	--	0.021	--	--	0.023

Total Nitrogen (TN; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
29-Cross Timbers	0	--	--	--	--	--	--	--
30-Edwards Plateau	2	7.04	--	--	7.13	--	--	7.21
32-Texas Blackland Prairies	3	1.55	--	--	2.04	--	--	2.17

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
29-Cross Timbers	0	--	--	--	--	--	--	--
30-Edwards Plateau	3	0.24	--	--	6.77	--	--	7.01
32-Texas Blackland Prairies	3	1.39	--	--	1.82	--	--	1.93

Phosphate (PO_4-P; mg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX
24-Chihuahuan Deserts	0	--	--	--	--	--	--	--
29-Cross Timbers	0	--	--	--	--	--	--	--
30-Edwards Plateau	2	0.025	--	--	0.0325	--	--	0.04
32-Texas Blackland Prairies	3	0.02	--	--	0.02	--	--	0.02

Chlorophyll-a (Chl-a; μg/L)

Level III Ecoregion	Count	MIN	10th	25th	Median	75th	90th	MAX

NA

Table A5.1.2. Frequency distribution by basin by level IV ecoregion of median nutrient concentrations in samples collected from springs in Texas, 2000-2010; these distributions are based on the reduced data with select monitoring types excluded.

Total Phosphorus (TP); mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24b-Chihuahuan Desert Grssland	0	--	--	--	--	--	--	--
29e-Limestone Cut Plain	0	--	--	--	--	--	--	--
30c-Balcones Canyonlands	3	0.05	--	--	0.06	--	--	0.06
32a-Northern Blackland Prairie	3	0.02	--	--	0.021	--	--	0.023

Total Nitrogen (TN; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24b-Chihuahuan Desert Grssland	0	--	--	--	--	--	--	--
29e-Limestone Cut Plain	0	--	--	--	--	--	--	--
30c-Balcones Canyonlands	2	7.04	--	--	7.13	--	--	7.21
32a-Northern Blackland Prairie	3	1.55	--	--	2.04	--	--	2.17

Nitrate plus Nitrite-Nitrogen (NO_X-N; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24b-Chihuahuan Desert Grssland	0	--	--	--	--	--	--	--
29e-Limestone Cut Plain	0	--	--	--	--	--	--	--
30c-Balcones Canyonlands	3	0.24	--	--	6.77	--	--	7.01
32a-Northern Blackland Prairie	3	1.39	--	--	1.82	--	--	1.93

Phosphate (PO_4-P; mg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX
24b-Chihuahuan Desert Grssland	0	--	--	--	--	--	--	--
29e-Limestone Cut Plain	0	--	--	--	--	--	--	--
30c-Balcones Canyonlands	2	0.025	--	--	0.0325	--	--	0.04
32a-Northern Blackland Prairie	3	0.02	--	--	0.02	--	--	0.02

Chlorophyll-a (Chl-a; μg/L)

Level IV	Count	MIN	10th	25th	Median	75th	90th	MAX

NA

Section 6: Historical Data Analysis

J.R. Chimka, E.M. Grantz, J.T. Scott, and B.E. Haggard

EXECUTIVE SUMMARY

Control charts are tools commonly used in industry or business to determine if a process is in a state of control that can be measured statistically. By using control charts, individual measurements can also be used to estimate error rates associated with water quality monitoring and to detect shifts in the population mean for a water quality parameter. In this study, this methodology was used to estimate values in Secchi transparency and nutrient concentrations that would be indicative of a change in water quality for 8-11 Texas stream and river, reservoir, and estuary sites, as well as the associated error rates for Type I and Type II errors. For inclusion in the analysis data belonging to a site-parameter combination had to meet 3 criteria: 1) the percentage of censored observations could not exceed 40%, 2) the data had to come from a lognormal or normal distribution, and 3) the population average had to be greater than 3 times the standard deviation. Possible control limits were defined as 1, 2, and 3x the standard deviation plus the mean, while a shift of interest indicating water quality degradation was 2x the mean for nutrient and chl-a concentrations and ½ the mean for Secchi transparency.

The results of this study indicated that traditional parametric control charts can be used to estimate control limits and run lengths for a number of Texas water bodies and parameters of interest using the currently available data. The largest number of site-parameter combinations meeting requirements for control chart analysis were among Texas streams and rivers. However, almost without exception across water body types, the water bodies for which traditional parametric control charts methods were well suited exhibited nutrient concentrations that were potentially above ecological thresholds for biological response on average. Limitations of traditional parametric control chart analysis, such as sensitivity to censored data and data distribution requirements were the likely cause for this finding.

These results indicate that traditional control chart procedures may be best suited for water bodies and parameters for which a large volume of high quality data collected using protocols with sensitive detection levels. Because the manager also makes decisions about shifts in parameters indicative of degradation and how far control limits should deviate from the mean, this method would also be best suited for water bodies where water quality management goals are specific and well-defined.

INTRODUCTION

Control charts are tools commonly used in industry or business to determine if a process is in a state of control that can be measured statistically. By using control charts, individual measurements can also be used to estimate error rates associated with water quality monitoring and to detect shifts in the population mean for a water quality parameter, if it can reasonably be assumed that a measurement belongs to a normally distributed dataset. In this study, this methodology was used to estimate values in Secchi transparency and nutrient concentrations that would be indicative of a change in water quality for 8-11 Texas stream and river, reservoir, and estuary sites, as well as the associated error rates for Type I and Type II errors.

Assuming normality in the distribution of the raw data, or in the data after undergoing a transformation such as natural logarithm, is necessary for these procedures, and the normal distribution assumption can be statistically tested in a number of straightforward ways, such as the W test for normality. If statistical tests confirm that the normal distribution is a reasonable assumption and good estimates of distribution parameters, such as the population average (m) and standard deviation (s), can be obtained, the probabilities of observing extreme values that may indicate a change in the average can be estimated. The normal distribution is illustrated Figure 6.1, with the greatest number of observations centered at the population average. The probability density, in turn, decreases as the value of an observation is increasingly greater or less in value than the average. The acceptance region, within which new measurements are assumed to belong to the original dataset distribution, is bounded by a lower control limit L and upper control limit U.

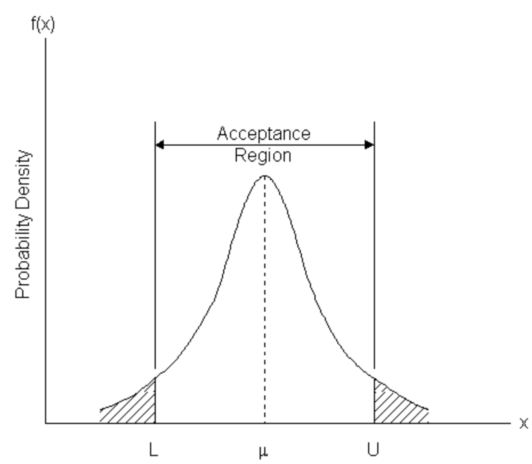


Figure 6.1. Normal distribution and acceptance region (Hendrickson 1998).

Table 6.1. Truth versus conclusions based on measurements of water quality.

Concentration Change	Conclude no increase	Conclude increase
Increase	Type II error	
No increase		Type I error

In turn, observations that are lower than L or higher than U are outside the acceptance region and have a low probability of belonging to the original distribution. Such measurements can be assumed to have come from a new distribution, with a different average but same shape and standard deviation. However it is important to acknowledge and compute the probabilities of Type I and II errors associated with these assumptions. For example, a Type I error would occur if it were concluded that concentration had increased when it had not in fact increased; a Type II error would occur if it were concluded that concentration had not increased when it had in fact increased (see Table 6.1).

Traditional control charts use so-called 3s limits to define U and L, where s is an estimate of the standard deviation. Therefore, it would be concluded that concentration had increased after observing a value more than three times the standard deviation greater or less than the average. For control charts assuming 3s limits (and the normal distribution), the probability of a Type I error is always p = 0.00135. Based on this probability of error, the expected number of measurements before a Type I error would occur can be calculated as the in-control average run length, or $ARL_0 = 1 / 0.00135 = 741$. Therefore, 741 sampling events would occur before a false conclusion that a change in water quality had occurred when it had in fact not occurred. The power to identify a change is then computed through identifying the probability of a Type II error, which occurs when it is concluded that concentration has not increased when it has in fact increased. Estimating Type II error requires that a shift of interest be defined. Tradeoffs exist between Type I and Type II errors, and traditional 3s limits may not accurately reflect the relative costs of these error types specifically associated with water quality monitoring. In the following example, details for applying a control chart approach to estimate Type II error and an examination of trade-offs between Types I and II error in setting control limits are outlined using TP concentrations at a Texas stream site.

METHODS

TCEQ provided lists of 10 stream and river, reservoir, and estuary segments of interest for historical data analysis. Raw data were compiled from all stations within the segments of interest from pivoted water quality raw databases described in Sections 1.1, 2.1, and 3.1 for streams and rivers, reservoirs, and estuaries, respectively. The focus parameters of this study were total phosphorus (TP; 00665), total nitrogen (TN; 00600C), Nitrate+nitrite-nitrogen (NO_x-N; 00630C), soluble reactive phosphorus (SRP; 00671C), spectrophotometric chlorophyll-a (chl-a spec; 32211), and fluorometric chlorophyll-a (chl-a fluoro; 70953).

For each segment and parameter combination, the following 3 criteria were required for subsequent control chart analysis: 1) the percentage of censored observations could not exceed 40%, 2) the data had to come from a lognormal or normal distribution, and 3) the population average had to be greater than 3 times the standard deviation. The appropriate level of censoring for inclusion in this analysis (i.e. <40%) was determined based on analyses discussed in Section 2.3 of this report. The data belonging to each site and parameter combination were tested for normality using the W test for normality in STATA statistical software both before and after undergoing natural logarithm transformation. If neither distribution assumption could be rejected (alpha = 0.05), the most appropriate distribution was assumed to be the one associated with the greater p-value. If both lognormal and normal distribution assumptions were rejected, that site and parameter combination was considered to not be in compliance with the second criteria for analysis outlined above.

Control chart process example

For Stream 1016-Greens Bayou, total phosphorus concentration measurements were obtained for the period January 1, 2000 to December 31, 2010. Results of the W test for normal data indicated that the normal distribution is reasonable for Greens Bayou TP concentrations (p = 0.6654). To build the control chart for n = 114 individual observations, the average and standard deviation of all TP measurements was calculated to be 1.46 mg/L and 0.51 mg/L, respectively. The traditional parameters used to detect a shift in the population to a new average in the positive direction (i.e. increase in concentration) is the upper control limit U = m + 3s. For TP at Stream 1016-Greens Bayou, U then equals 3.01. Since TP concentrations for Stream 1016 conform to the normal distribution and a 3s limit was established, the probability of measuring a TP concentration greater than 3.01 while the population mean remains unchanged is p = 0.00135. The expected number of measurements before a Type I error, or reaching the false conclusion that the average concentration has increased when it has not, is then ARL_0 = 1 / 0.00135 = 741. In other words, 741 sampling events would occur before a Type I error.

Calculating Type II error requires defining a shift of interest. For nutrient and chl-a concentrations, the shift-of-interest was defined as 2m, or 2 times the mean. For parameters with concentrations, the upper control limit U is active, meaning that exceeding U would indicate worsening water quality. For Secchi transparency, the shift-of-interest was defined as m / 2, or half the mean, and lower control limit L is active, meaning measurements less than L would indicate worsening water quality. For the lognormal distribution assumption concentration shift-of-interest is from exp [m (logarithms)] to exp [m (logarithms)] + exp [m (logarithms)]. For the lognormal distribution assumption depth shift-of-interest is from exp [m (logarithms)] to exp [m (logarithms)] / 2. Results for Type II error probabilities and run lengths are all gauged to this definition of a shift of interest, therefore, these results would change if shift-of-interest were set using a different definition.

Therefore, for the example stream 1016-Greens Bayou, for which TP concentrations conformed to the normal distribution, the average TP concentration shift-of-interest was defined to be from m = 1.46 mg/L to 2m = 2 (1.46 mg/L) = 2.9 mg/L. The probability of Greens Bayou TP measuring less than 3.01 given TP concentrations are normally distributed (2.9, 0.51^2) is 0.56. Conversely, the probability of a Type II error is 0.65. The number of measurements needed in order to detect the positive shift-of-interest is the out-of-control run length, or ARL_1 = 1 / (1 – 0.56) = 2.3. In other words, a 2m shift of interest in Green Bayou average TP concentration could be detected after 3 sampling events, when the upper control limit U is set at 3s.

Type I and II error rates are directly related and minimizing one type of error can increase the other type. To better understand these tradeoffs, see Table 6.2 based on Greens Bayou TP concentrations and the positive shift-of-interest defined as from m = 1.46 to 2m = 2.92. Columns are 1s, 2s and 3s upper control limits U, probability of a Type I error P (Type I), in-control average run length ARL0 between Type I errors, probability of a Type II error P (Type II), and out-of-control average run length ARL1 to detect a true increase in Greens Bayou TP concentration. Results of Type I error are general, for any normal data, while results of Type II error are specific to the parameters estimated for Greens Bayou TP concentrations.

Table 6.2. Consequences of setting different control limits on Types I and II error rates (p) and run lengths (ARL). Error rates and run lengths for Type I error are generalized, but estimates for Type II error parameters are specific to Greens Bayou TP concentrations and the shift-of-interest defined in this study.

s limit	U	P (Type I)	Sample # between Type I errors	P (Type II)	Sample # until change is detected
1	2.0	0.16	6	0.03	1
2	2.5	0.023	44	0.20	2
3	3.0	0.0014	741	0.56	3

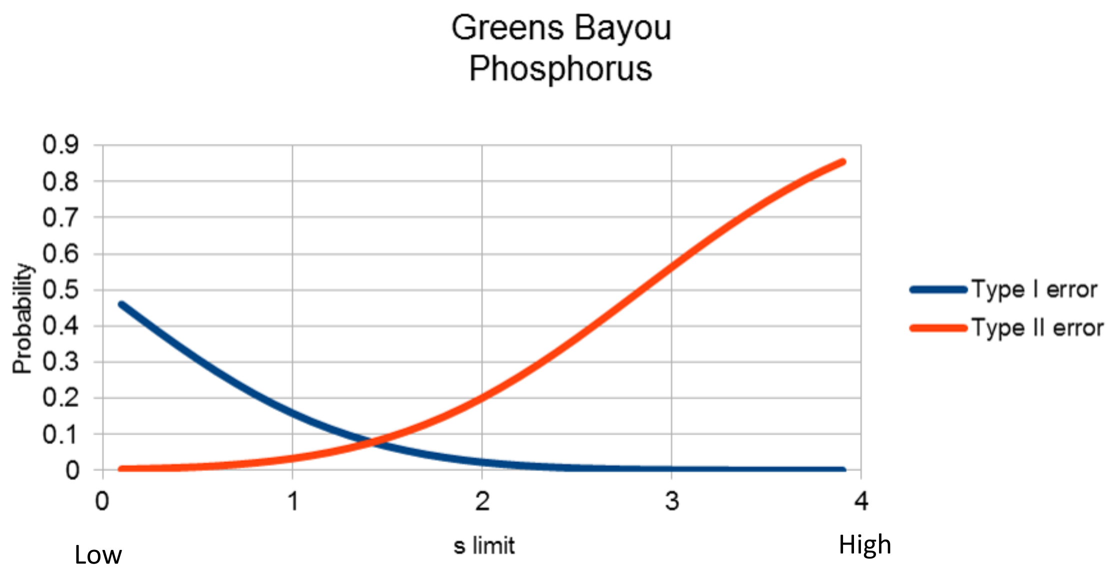


Figure 6.2. Type I and Type II error probabilities for different possible s limits for TP concentrations at Stream 1016-Greens Bayou. The number of s limits denotes the number of times the population standard deviation has been added to the population mean in order to define a control limit. For parametric control charts, 3s limits are traditional, but not required, and this example illustrates the fact that setting different s limits adjusts the probabilities of Types I and II errors.

This tradeoff between Type I and II error can also be illustrated graphically (Figure 6.2). For Greens Bayou TP concentration, error probabilities are equal (p~0.10) when upper control limit U is approximately equal to 1.5s. Therefore, assuming equal cost associated with Type I and II errors, U=1.5s would be the optimal control limit. In another scenario, if Type II error were considered more costly and were more desirable to minimize, the upper control limit U should be less than 1.5s. In contrast, if Type I error were considered more costly and were more desirable to minimize, the upper control limit should be greater than 1.5s. Every segment and parameter combination that met the 3 criteria outlined previously were subjected to the analysis outlined above for Stream 1016-Greens Bayou TP concentrations. Results for the remaining segment-parameter combinations are summarized following.

RESULTS AND DISCUSSION

Streams

Historical data were provided for 11 streams for the 7 parameters of interest. Of these 77 possible combinations 6 contained no data or were determined to have possible missing data or metadata. Of the remaining 71 segment-parameter combinations, 25 include censored data, and 18 of these 25 contained data that were greater than 40% censored. Therefore 53 datasets were tested for lognormal and normal distribution assumptions. Out of 53 datasets available for testing, 34 could reasonably be assumed to be lognormally and / or normally distributed. Specifically, 21 of 34 were assumed to be lognormal, and 13 of 34 were assumed to be normal.

Control limits and out-of-control run lengths (Type II error) for TP, TN, NOX-N, SRP, and chl-a spec and fluoro concentrations, as well as Secchi transparency, for Texas streams are summarized in Table 6.3. Refer to Table 6.2 for the Type I error probabilities and in-control run lengths associated with each U (nutrient or chlorophyll-a concentrations) or L (Secchi transparency), i.e. 1s, 2s. 3s. All estimates of the out-of-control run length were rounded up to the next greatest whole number to approximate the number of sampling events that would be required before a change was detected at a given control limit.

Table 6.3. Control limits and out-of-control run lengths for TP, TN, NO_x-N, SRP, Secchi transparency, and chlorophyll-a (chl-a) measured fluorometrically and spectrophotometrically for Texas streams and rivers. Sample size is denoted by "n." Under normal, "0" indicates lognormally distributed, while "1" indicates the normal distribution.

Total Phosphorus (mg/L)

Stream	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
204	39	0	0.47	2	0.94	7	1.9	44
1016	114	1	2.0	1	2.5	2	3.0	3
1202	19	0	0.41	2	0.62	3	0.96	12
1428	65	0	0.96	2	2.1	8	4.5	57
1912	33	0	3.32	2	5.0	3	7.4	10

Total Nitrogen (mg/L)

Stream	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
204	22	0	2.3	2	3.4	3	5.2	11

402	38	0	0.86	1	1.1	2	1.4	2
1016	37	1	8.6	1	11	2	13	2
1017	39	1	8.0	2	11	2	13	5
1202	18	0	1.9	2	2.8	3	4.1	10
1247	12	1	11	2	15	2	19	5
1428	60	0	5.4	2	9.9	5	18	32

Nitrate+nitrite-Nitrogen (mg/L)

Stream	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
1016	39	1	7.1	1	9.0	2	11	3
1017	40	1	6.8	2	9.3	3	12	10
1247	42	0	14	2	8	7	57	47
1428	60	0	4.7	2	9.5	7	19	44
2112	42	0	1.5	2	2.2	3	3	10

Orthophosphate (mg/L)

Stream	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
402	40	1	0.06	1	0.08	1	0.09	2
1016	62	1	2.0	1	2.5	2	3.0	3
1017	60	0	1.4	2	2.4	4	4.0	22

1202	20	0	0.27	2	0.46	4	0.77	22
1428	65	0	0.92	2	2.2	9	5.1	71
1912	32	1	3.0	1	3.8	2	4.6	3

Chorophyll-a (µg/L)

Stream	n	Normal	1s (µg/L)	Sample # until change is detected	2s (µg/L)	Sample # until change is detected	3s (µg/L)	Sample # until change is detected
204 spec	7	0	70	3	150	7	310	52
204 fluoro	16	0	66	3	150	8	320	59
1017 fluoro	14	0	13	2	25	6	47	35
1202 fluoro	9	1	3.4	1	4.4	2	5.4	3
1247 spec	6	0	16	3	57	14	200	139
1428 spec	44	0	5.9	3	14	9	33	71
3211 fluoro	18	0	3.1	3	50	8	110	60

Secchi transparency (m)

Stream	n	Normal	1s (m)	Sample # until change is detected	2s (m)	Sample # until change is detected	3s (m)	Sample # until change is detected
805	93	0	0.07	2	0.03	6	0.02	41
1202	24	1	0.10	3				
1428	66	1	0.25	2	0.09	5		
1912	30	1	0.56	2	0.08	6		

Reservoirs

Historical data were provided for 10 reservoirs for the 7 parameters of interest. Of these 70 possible segment-parameter combinations, 5 did not contain any data. Of these 65 remaining datasets, 44 included censored data. Every segment contained censored NO_x-N, SRP, and chl-spec data. Of the censored datasets, 29 had greater than 40% censored data. Therefore, 36 datasets remained to test for lognormal and normal distribution assumptions. Out of 36 datasets available for testing, 21 could be reasonably assumed to fit a lognormal and / or normal distribution. For every segment, NO_x-N and SRP data could not be assumed to fit a (log)normal distribution. Data for 14 of the segment-parameter combinations were assumed to be lognormal, and 7 were assumed to be normal.

Control limits and out-of-control run lengths (Type II error) for TP and chl-a spec and fluoro concentrations for Texas reservoirs are summarized in Table 6.4. Refer to Table 6.2 for the Type I error probabilities and in-control run lengths associated with each U (nutrient or chlorophyll-a concentrations) or L (Secchi transparency) level (i.e. 1s, 2s. 3s).

Table 6.4. Control limits and out-of-control run lengths for total phosphorus (TP), chl-a (both spec and fluoro), and Secchi transparency for Texas reservoirs. Sample size is denoted by "n." Under normal, "0" indicates lognormally distributed, while "1" indicates the normal distribution.

Total Phosphorus (mg/L)

Reservoir	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
1002	169	0	0.21	1	0.27	2	0.36	3

Chlorophyll-a (µg/L)

Reservoir	n	Normal	1s (µg/L)	Sample # until change is detected	2s (µg/L)	Sample # until change is detected	3s (µg/L)	Sample # until change is detected
403 fluoro	15	1	28	1	37	2	45	4
1002 fluoro	18	1	49	2	67	3	86	13
1012 spec	26	1	26	2	35	2	43	5

1203 fluoro	33	0	24	2	48	7	97	45
1404 spec	45	0	6	2	12	7	24	47
1408 spec	45	1	14	2	19	3	24	10
2303 spec	32	0	23	2	47	7	97	49

Secchi transparency (m)

Reservoir	n	Normal	1s (m)	Sample # until change is detected	2s (m)	Sample # until change is detected	3s (m)	Sample # until change is detected
403	41	1	0.91	1	0.76	1	0.62	2
507	157	1	0.70	2	0.44	3	0.18	9
1002	39	0	0.20	2	0.12	4	0.08	16
1012	61	0	0.61	2	0.43	2	0.31	7
1203	63	0	1.1	1	0.81	2	0.63	3
1220	70	0	1.6	2	1.1	2	0.78	5
1404	66	0	2.0	2	1.3	3	0.81	13

Estuaries

Historical data were provided for 8 estuaries for the 7 parameters of interest. Data were available for all of the 56 segment-parameter combinations, but 27 included censored data. All segments contained censored chl-a data (both methods). For estuaries 13 segment-parameter combinations had greater than 40% censored data; therefore, 43 datasets remained to test for lognormal and normal distribution assumptions. Out of 43 datasets available for testing, 18 could be reasonably assumed to have a lognormal and / or normal distribution. It was not possible to fit a (log)normal distribution to any of the NO_x-N or chl-a spec datasets. Data for 16 of the segment-parameter combinations were assumed to be lognormal, and 2 were assumed to be normal.

Control limits and out-of-control run lengths (Type II error) for TP, TN, SRP, and chl-a fluoro concentrations, as well as Secchi transparency, for Texas estuaries are summarized in Table 6.5. Refer to Table 6.2 for the Type I error probabilities and in-control run lengths associated with each U (nutrient or chlorophyll-a concentrations) or L (Secchi transparency) level (i.e. 1s, 2s. 3s).

Table 6.5. Control limits and out-of-control run lengths for TP, TN, SRP, chl-a fluoro and Secchi transparency for Texas estuaries. Sample size is denoted by "n." Under normal, "0" indicates lognormally distributed, while "1" indicates the normal distribution.

Total Phosphorus (mg/L)

Estuary	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
2422	43	0	0.24	2	0.33	2	0.47	7
2452	21	0	0.15	2	0.24	3	0.36	13
2472	68	0	0.16	2	0.24	3	0.36	11

Total Nitrogen (mg/L)

Estuary	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
2422	42	0	1.5	2	2.4	4	3.8	15
2451	92	0	1.5	2	2.6	5	4.6	27
2452	20	0	1.2	2	1.9	3	2.9	12

Orthophosphate (mg/L)

Estuary	n	Normal	1s (mg/L)	Sample # until change is detected	2s (mg/L)	Sample # until change is detected	3s (mg/L)	Sample # until change is detected
2422	39	0	0.28	3	0.62	8	1.4	60

Secchi transparency (m)

Estuary	n	Normal	1s (m)	Sample # until change is detected	2s (m)	Sample # until change is detected	3s (m)	Sample # until change is detected
2412	85	0	0.04	3	0.02	9	0.01	66
2422	39	0	0.24	3	0.11	7	0.05	52
2454	23	1	1.4	2				
2472	67	0	0.28	2	0.15	6	0.08	33
2481	246	0	0.61	2	0.40	3	0.26	13
2491	229	0	0.40	2	0.24	4	0.14	**21**

Chlorophyll-a (µg/L)

Estuary	n	Normal	1s (µg/L)	Sample # until change is detected	2s (µg/L)	Sample # until change is detected	3s (µg/L)	Sample # until change is detected
2412	31	0	8.8	2	13	3	21	12
2422	14	0	18	2	29	4	47	19
2451	48	1	13	2	16	3	21	8
2452	14	0	16	2	33	7	66	45
2456	17	0	56	2	110	7	220	44

Potential Uses for a Traditional Parametric Control Chart Method

Where the distribution and data quality requirements outlined in the introduction are met, traditional parametric control charts could be a useful monitoring tool as part of a “net zero change” water quality management strategy in Texas water bodies. These methods allow for error estimates associated with the monitoring procedures and are well-suited for a resources-limited management situation in which Type II error is likely more costly than that Type I. Furthermore, these analyses can be used to estimate the number of samples required to detect change under a variety of scenarios, providing information on the sampling intensity required if a time frame within which detection of change should occur is known. Because these procedures require managers to decide on a shift of interest and how far control limits should deviate from current means, this tool would be most effective for water bodies for which specific targets for maintaining water quality are established, as well as a shift in average conditions that would signal degradation. A large quantity of high quality data are needed for control chart procedures to produce good estimates of control limits and run lengths; therefore, this tool would be most effective for sites that have undergone intensive sampling regimes with sensitive detection levels for parameters of interest.

The results of this study indicated that traditional parametric control charts can be used to estimate control limits and run lengths for a number of Texas water bodies and parameters of interest using the currently available data. The largest number of site-parameter combinations meeting requirements for control chart analysis were among the selected Texas streams and rivers. Control limits and run lengths were estimated for all parameters of interest at a minimum of four stream segments. For streams and rivers, Secchi transparency control limits were typically limited to 1s or 2s. From a theoretical perspective, this occurred because the lower control limit is active for Secchi transparency, and this variable often conformed to a normal distribution. Therefore, though it is only possible to use Secchi transparency as an indicator of worsening water quality to a lower limit of surface visibility, the control chart procedures estimated negative values as control limits, which is impossible in practice. The number of site-parameter combinations for which control limits and run lengths could be estimated was more restricted for reservoirs and estuaries. However, Secchi transparency appeared to be a more effective metric for tracking changes in water quality for these water body types.

Almost without exception, across water body types, the water bodies for which traditional parametric control charts methods were well suited exhibited nutrient concentrations that were potentially above ecological thresholds for biological response on average. One likely reason for this finding is that characteristics of the available data for lower trophic water bodies caused the data not to meet requirements of the analysis. Data censorship was a common problem for these sites. Systematic variability in some parameters related to environmental controls could also interfere with the assumptions required for control chart procedures, such as the seasonal variability in dissolved nutrients and chl-a concentrations that is frequently observed in lakes, reservoirs, and estuaries. For example, with only one exception, the data associated with all site-parameter combinations for reservoirs and estuaries involving dissolved nutrients failed to conform to a (log)normal distribution in this study. For these water body types, chl-a data often did conform to a log(normal) distribution, but the control limits estimated based on standard deviations often resulted in extremely high estimates for these limits that may not be practical for real-world management. These possible limitations to the use of control charts, and others, will be discussed in greater detail in the section following.

Limitations of the Traditional Parametric Control Chart Method

Distribution assumptions. The traditional parametric control chart method relies on the assumption that a dataset is (log)normally distributed. In many cases, that was not a valid assumption for nutrient and chl-a concentration and Secchi transparency data from Texas streams and rivers, estuaries and reservoirs. This was particularly true of dissolved nutrient data from estuaries and reservoirs, which may reflect the fact that dissolved nutrient fractions can be highly seasonably variable in these types of systems.

Analyses relying on other possible distribution assumptions are available, but, at present, our understanding of their usefulness for these data is limited. Where neither the lognormal nor normal distribution should be assumed, investigation of other distribution families may be interesting future work. Bimodal distributions may require regression analysis to identify an external variable which divides measurements into separate populations. One hypothesis would be that average water quality conditions and distributions might differ among stations within a segment. This hypothesis could be tested with currently available data. This issue could also be addressed through future work by exploring non-parametric control chart methods, which do not require a distribution assumption.

Censored datasets. Another practical limitation of this analysis was the presence of left- and right-censored data in the water quality dataset. Censored data are problematic for assuming distributions and estimating distribution parameters. The control chart method only functions when good assumptions about data distribution and associated measures of central tendency can be made. It is potentially questionable to suggest that a dataset's distribution can be known when very little information about the value of some or many of the individual observations is available.

The presence of censored data raises additional research questions. To what extent is it possible to detect changes in water quality from historical conditions when historical data are censored? Analyses presented in Section 2.3 indicated that datasets with greater than approximately 40% censored data required analysis with techniques specifically developed for estimating measures of central tendency in censored datasets in order to more accurately estimate those metrics. These findings were used to set guidelines for which segment-parameter combination datasets were potentially suitable for control chart analysis in this study, but the effects of censored datasets on control chart results have not been tested as was done for medians, means, and standard deviations in Section 2.3.

In order to minimize the effects of censored data, future work could include a generalization of Tobit regression called interval regression. Ranking procedures such as Kaplan-Meier survival analysis (see Section 2.3) could also be used to estimate the upper percentiles in censored datasets as possible control limits. Though statistical methods are available that can extract information from censored datasets, analyses in Section 2.4 indicated that ecologically relevant thresholds, especially for TP concentrations in Texas reservoirs, may be lower than common detection limits. Determining control limits, thresholds, and other metrics with greater confidence at a future time may require moving to analytical protocols with more sensitive detection limits for future data collection efforts.

Type I and II error tradeoffs. While Type I error rates and in-control run lengths associated with defined upper and lower confidence limits are standard, upper and lower confidence levels can be adjusted to optimize error rates and run lengths in both Type I and II error. Furthermore, Type II error is a function of a defined shift of interest. In theory, it could be possible and preferable to select control limits to optimize error rates and run-lengths according to a distribution assumption, relative costs of Type I and Type II errors, and shift-of-interest (as in Figure 6.2). This course of action would be relatively complicated and computationally intensive, but could be worth exploring through future work, especially if the relative costs of Type I and II error are known. For example, a statement about relative costs might be: “The cost to conclude concentration has increased, when it has not in fact increased (Type I error), is twice as great as the cost to conclude concentration has not increased, when it has in fact increased (Type II error).”

REFERENCES

Hendrickson, Chris (1998), Project Management for Construction, http://pmbook.ce.cmu.edu