12-2008

Estimating Total Phosphorus and Total Suspended Solids Loads from High Frequency Data

Amber Spackman Jones
Utah State University

ESTIMATING TOTAL PHOSPHORUS AND TOTAL SUSPENDED SOLIDS LOADS FROM HIGH FREQUENCY DATA

by

Amber Spackman Jones

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

Approved:

_____	_____
David K. Stevens Major Professor	Bethany T. Neilson Committee Member
_____	_____
Nancy O. Mesner Committee Member	Byron R. Burnham Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2008

ABSTRACT

Estimating Total Phosphorus and Total Suspended Solids

from High Frequency Data

by

Amber Spackman Jones, Master of Science

Utah State University, 2008

Major Professor: Dr. David K. Stevens
Department: Civil and Environmental Engineering

Frequently measured turbidity was examined as a surrogate for total phosphorus (TP) and total suspended solids (TSS) loads at two locations in the Little Bear River, Utah, USA. Using regression techniques, equations were developed for TP and TSS as functions of turbidity. The equations accounted for censored data, and additional explanatory variables to represent hydrological conditions were considered for inclusion in the equations. By using the resulting surrogate relationships with high frequency turbidity measurements, high frequency estimates of TP and TSS concentrations were calculated. To examine the effect of sampling frequency, reference loads were determined from the concentration records for two water years. The concentration records were artificially decimated to represent various frequencies of manual grab sampling from which annual loads were calculated and compared to the reference loads.

(127 pages)

ACKNOWLEDGMENTS

I thank my advisor, Dr. David K. Stevens, for sharing his expertise and opinions. I am grateful for his patience. I also appreciate the contributions of my other committee members, Nancy O. Mesner, who provided guidance with ideas and data, and Dr. Bethany T. Neilson, who provided insight into the research process and would not let me quit. Many thanks to my colleague Jeff Horsburgh for being the brains behind the sensor network, for his programming skills, and for his willingness to help me work through numerous questions. Thank you to all who assisted in site maintenance and the collection and analysis of samples, including Sandra Guerrero, Emily Saad, Ian Washbourne, Michael Stevens, Eric Peterson, and Su Anderson. I am grateful to the landowners along the Little Bear River who generously granted us access to their land. I also thank the National Science Foundation and the US Department of Agriculture for providing funding for this research.

I am indebted to my family for their confidence in me and for their encouragement, my dear friend Kaylynn who accompanied me through this adventure, and especially to my husband, Tanner, who has been subject to much of this process. I thank you for your time, attention, and love.

Amber Spackman Jones

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Traditional water quality monitoring programs typically rely on the analysis of grab samples, and the frequency of sample collection is dependent on limitations in resources such as personnel, budget for sample analysis, and supplies. As a result, sample collection often happens on a weekly to monthly basis, which, depending on the variable and the location, may not provide an adequate representation of the behavior of most water quality constituents. Concentrations of many water quality constituents can vary at scales of minutes to hours. In general, even if resources were unlimited, it would be logistically infeasible to collect samples at these high frequencies over extended time periods. In order to estimate constituent transport over time, grab sampled concentration data are paired with discharge (often measured more frequently than concentration) to calculate loads. If sampling is conducted infrequently, these estimates may grossly over or under estimate the true constituent loads. An additional drawback is that a complicated calculation method is often necessary to estimate loads from infrequently collected concentration data to account for averaging over those periods and for discharge collected more frequently than concentration.

For some constituents, in situ sensors can be used for high frequency monitoring returning large datasets over relatively long time periods. Variables commonly measured in situ include physical parameters such as water level, pH, specific conductance, dissolved oxygen, and turbidity. Additionally, UV-VIS spectroscopy and ion-specific sensors can be used in situ to quantify constituents such as nitrate, nitrite, chlorophyll, and chemical oxygen demand. Despite developments in sensor technology, there are still

important water quality constituents that cannot practically be measured in situ or in real time over an extended period of time. For example, total phosphorus samples must be digested and analyzed in the lab. Consequently, the number of available measurements is limited in most watersheds. As total phosphorus is often associated with particulates including soils, animal waste, and vegetation, its loading, along with total suspended sediment loading, is likely to increase during storm events and times of high runoff when erosion occurs. These are periods when representative grab sampling can be especially difficult and may not often be conducted. As an alternative to grab sampling, high frequency, in situ measurements can be used as surrogates to estimate properties such as contaminant concentration. A common surrogate measure is turbidity, an optical measure of the scattering of light passing through a sample of water due to colloidal and suspended matter.

This research examines turbidity as a surrogate measure for total phosphorus (TP) and total suspended solids (TSS) on the Little Bear River, Utah, USA and uses the results to assess the effects of sampling frequency on load calculations. Historically, TP and TSS have been constituents of concern on the Little Bear River, and the Utah Department of Environmental Quality (DEQ) included the river on its 303(d) list of impaired water bodies for total phosphorus. Additionally, the Little Bear is one of 11 test beds in the Water and Environmental Research (WATERS) Network designated to research environmental observatory design and methods for improved understanding of instream processes including high frequency data collection and surrogate measures.

Using high frequency turbidity data and intermittently sampled TP and TSS, equations were developed to estimate TP and TSS as functions of turbidity from which high frequency estimates of concentration were generated. The relationships account for censored data, and additional categorical variables representing hydrological conditions were considered. The continuous concentration estimates, used in conjunction with high frequency discharge data, were used to calculate annual loads. In order to examine the effect of sampling frequency on load estimates, the continuous concentration and discharge series were artificially decimated to represent periodic, less frequent grab sampling. This subsampling was conducted at hourly, daily, weekly, and monthly frequencies, from which annual loads were calculated. Multiple realizations of daily, weekly, and monthly sampling were generated by randomizing the selection of concentration and discharge values. The results were compared to the reference loads calculated from the high frequency discharge and concentration data. Additionally, consistently sampling at the same time of the day and the same day of the week were examined in order to examine the effect of timing of sample collection on load estimates.

Chapter 2 provides a review of literature establishing a background for this research and descriptions of related work. Chapter 3 describes the procedures used to develop the surrogate relationships for TSS and TP at both locations. The continuous concentration datasets and reference loads are presented in Chapter 4, which also examines the loads calculated by subsampling at different frequencies. Chapter 5 summarizes the results of the analyses, Chapter 6 details the engineering significance of this research, and Chapter 7 suggests topics of future research stemming from this work.

CHAPTER 2

LITERATURE REVIEW

2.1 Study area: Little Bear River

The site of this research is the Little Bear River in northern Utah, USA, which drains a semi-arid watershed with hydrologic behavior dominated by spring snowmelt runoff. The Little Bear River watershed encompasses an area of approximately 740 km^2, the headwaters are in the Bear River Mountain Range, and elevations range from 1340 m to 2700 m. The river has two principal subdrainages, the East Fork and the South Fork. There are two reservoirs within the drainage: one in the upper watershed on the East Fork (Porcupine Reservoir) and another in the lower watershed (Hyrum Reservoir). Both reservoirs are operated by canal companies and store water for the summer irrigation season. Below the reservoirs and at other locations along the river are agricultural diversions that greatly influence the hydrology of the Little Bear River. The land use within the watershed is primarily agricultural with a general distribution of 70 percent grazing land and forest, 19 percent irrigated cropland, and 7 percent dry cropland. There are a number of small towns within the watershed, and the area has exhibited population growth of 32 percent between 1990 and 2000 (US Census Bureau, 2000).

The geologic material surrounding and underlying the Little Bear River is primarily limestone and dolomite rocks (Schaefer et al., 2006). In the upper watershed, most of the underlying soils consist of high slope (30-50 percent) silty 4olluviums deposits, and the depth to the water table is generally greater than 2 meters. In contrast, in the lower watershed, the soils are primarily loamy lacustrine deposits of low slopes (0-

3 percent) with a depth to the water table of 0.75-1.5 meters or less (Soil Survey Staff, 2008)

The Little Bear drains into an arm of Cutler Reservoir, a shallow eutrophic reservoir on the Bear River, a tributary to the Great Salt Lake. Cutler Reservoir has been listed as impaired with respect to low dissolved oxygen concentrations driven by algae growth due to high phosphorus levels (Utah DEQ, 2006b). Consequently, a Total Maximum Daily Load (TMDL) is currently under development for Cutler Reservoir. TMDLs have already been developed on many of the reservoir's tributaries, including the Little Bear River. The State of Utah has applied a guideline of 0.05 mg/L for maximum instream total phosphorus concentrations, which has not been met in the Little Bear River (Utah DEQ, 2000a, 2006a), and a TMDL was completed in 2000 (Utah DEQ, 2000b). A TMDL was also completed for Hyrum Reservoir, which often has algal blooms, that requires an in-lake total phosphorus concentration to meet an endpoint of 0.025 mg/L (Utah DEQ, 2002). According to the TMDL studies, the reduction in phosphorus loading must be achieved through best management practices implemented by landowners and community members.

2.2 Project funding and context

To address deficiencies in the current state of understanding of hydrologic systems, the Water and Environmental Research Systems (WATERS) Network was created consisting of 11 environmental observatory test bed sites across the United States. The test beds are examining techniques and technologies for larger scale environmental observatory design and operation. Research topics include innovative methods for

constituent estimation, deployment of environmental sensor networks, development of modeling tools, and standardization of data storage and publication (WATERS Network, 2006; Montgomery et al., 2007). The Little Bear River was selected as a test bed site with the following objectives: 1. Develop an integrated monitoring system of data collection and surrogate measurements, 2. Assess high frequency nutrient loading in relation to flow regime, watershed characteristics, and management practices, and 3. Develop two-way linkages between field sensors and a central database including modeling tools or software for data access and watershed management (Utah Water Research Laboratory, 2007). Additional funding was provided by the United States Department of Agriculture through the Conservation Effects Assessment Program, a national study evaluating the results of conservation practices implemented by private landowners.

2.3 Phosphorus and suspended solids

Phosphorus is an essential nutrient in aquatic systems as it is required for most forms of life. However, over-enrichment of phosphorus in water bodies can cause increased algal growth leading to eutrophication in lakes and reservoirs and excessive periphyton growth in rivers (Hem, 1985; US EPA, 1986; Mueller and Helsel, 1996). Concerns with eutrophic water bodies include aesthetics for natural waters and drinking water sources as well as reduced dissolved oxygen levels, which adversely affect fish and other forms of aquatic life. Phosphorus is found naturally in some soils, but significant amounts may also be contributed to aquatic systems by anthropogenic sources such as fertilized fields, animal waste, wastewater treatment plants, and industries that use

phosphorus in cleaning processes (Hem, 1985; Mueller and Helsel, 1996). Depending on the source, phosphorus is frequently associated with suspended sediments, which may also be a water quality concern (Kronvang et al., 1997; Heimlich, 2003). Not only do suspended sediments transport contaminants such as nutrients, pesticides, and metals, high levels of suspended sediment can be detrimental to aquatic life, decrease the recreational quality of a water body, complicate water treatment, and interfere with the operation of hydraulic structures (US EPA, 1986).

2.4 Water quality monitoring

Literature regarding the design of water quality sampling programs and monitoring networks is widely available (Ward et al., 1990; Harmancioglu et al., 1999), and networks have been established at varying scales. On a national scale, for example, the National Stream Quality Accounting Network (NASQAN) was implemented in 1974 by the United States Geological Survey (USGS) to study the water quality of the nation's five largest rivers (Mississippi, Rio Grande, Yukon, Colorado, and Columbia). There are only a few stations on each river that are generally sampled 5 to 15 times annually. The program examines chemical and sediment transport on relatively large scales, both temporally and spatially (Hooper et al., 1997). Sampling programs are often initiated on a smaller watershed scale to meet various objectives such as assessment of the effectiveness of management practices, providing data for modeling efforts, and for determination of compliance with water quality standards (Oblinger, 2004). These traditional water quality monitoring programs rely on grab samples that typically are collected with a frequency too low to fully characterize the range in ambient

concentrations and to accurately calculate loads of water quality constituents over time (Ferguson, 1987; de Vries and Klavers, 1994; Coynel et al., 2004; Etchells et al., 2005; Johnes, 2007).

Traditional grab sampling at weekly or monthly intervals often misses storm events, periods when loading of solids, nutrients, and bacteria are increased due to non-point source runoff. Croke and Jakeman (2001) discuss streams with rapid hydrological response that are especially subject to erosion resulting in increased transport of sediment and associated nutrients during storm events. Nolan et al. (1995) showed that increases in concentration of the various species of phosphorus were closely associated with the occurrence of precipitation. Kronvang et al. (1997) compared intensive storm sampling with fortnightly sampling and found that the infrequent sampling significantly underestimated the transport of sediments and phosphorus. Gray and Glysson (2002) make the generalization that approximately 90 percent of the sediment transport in smaller streams occurs in 10 percent of the time.

2.5 High frequency monitoring

Important periods in constituent transport are usually missed or underrepresented by traditional grab sampling (Richards and Holloway, 1987; Christensen et al., 2002; Jordan et al., 2007), but can be captured by high frequency water quality monitoring. Continuous, high frequency monitoring also elucidates seasonal and diurnal trends that may be overlooked by traditional periodic grab sampling (Grayson et al., 1997; Christensen, 2001; Tomlinson and De Carlo, 2003; Kirchner et al., 2004; Scholefield et al., 2005). Automating high frequency monitoring reduces the logistics and personnel

required for grab sampling that is representative (Grayson et al., 1997), minimizes errors in transcription and improves the turnaround between the collection and the use of field data (Vivoni and Camilli, 2003), and provides data at increased temporal and spatial scales for extended time periods (Kirchner et al., 2004).

Tomlinson and De Carlo (2003) collected high frequency water quality data on three Hawaiian streams, illustrating patterns that could not be captured by sampling monthly, weekly, or even daily. Some of their findings included multiple discharge peaks within 24-hour periods, 60-fold increases in turbidity within 15 minutes, and 30-fold increases of turbidity within 5 minutes. Their data collection also demonstrated cyclical fluctuations in temperature, pH, and dissolved oxygen and helped reveal tidal influence.

Kirchner et al. (2004) assert the importance of collecting high frequency data over extended time periods. The authors make the analogy that drawing conclusions based on infrequent measurements is like looking through a blurry telescope where only the most prominent features of the watershed are visible. On the other hand, if intensive sampling is conducted only during certain events, it is like viewing the watershed through a pinhole where fine details are visible, but the entire picture is obscured.

Pressl et al. (2004) describe benefits of automated water quality measurements including the ability for quick action in response to negative water quality changes, a reduction in overall monitoring costs, and higher resolution data for better identification of trends. The authors conducted a study employing automated monitoring using one station with real time, in situ equipment. Challenges encountered in this study included

the need for in situ calibration, river stratification, low water levels, and faulty probes, but overall, the data collection was deemed successful.

High frequency data collection is enhanced by the real time acquisition of water quality data. Vivoni and Richards (2005) describe the benefit of closely linking data collected in real time with a water quality model that can be run continuously. The results of model simulations can be used to better direct sampling, and the field data can be used to more frequently refine model parameters and results.

Additionally, when water quality criteria are exceeded, real time data allows immediate action to be taken. Christensen et al. (2002) explain that if violations of bacteriological criteria are identified in real time, managers can act to insure that human and animal contact with the water is prevented. Real time control of drinking water sources and the prevention of eutrophication can also be facilitated by a more rapid response. Fleischmann et al. (2002) installed a real time sensor network to serve as a warning system for drinking source water protection, which is accessed and remotely controlled through a web interface.

In situ sensors have been installed in sewer systems in order to consistently quantify pollutant loads discharged from combined sewer networks. Parameters commonly measured are water level, ammonia, nitrate, pH, conductivity, carbonaceous oxygen demand, total organic carbon, and dissolved and suspended solids. In the system described by Gruber et al. (2005), observations were made every three minutes with more frequent (one-minute) observations made when the channel water level exceeded a threshold corresponding to overflow conditions. Most of the monitoring methods

employed in this study were based on UV-VIS-absorption, which requires frequent calibration. Vanrolleghem et al. (2005) describe a system of sensors in receiving water bodies that used real time, in situ data to trigger flow controls within the wastewater treatment system if pollutant levels were exceeded.

In situ sensors are commonly used for physical parameters such as water level and temperature and some water quality constituents such as pH, conductivity, dissolved oxygen, and turbidity. More recent technological advances allow the measurement of some chemical species such as the ions of metals and nutrients. As mentioned, UV-VIS spectroscopy has been used to measure constituents such as nitrate, nitrite, chlorophyll, and chemical oxygen demand (Fleischmann et al., 2002; Pressl et al., 2004; Gruber et al., 2005), and Winkler et al. (2004) describe the use of ion-sensitive sensors for the real time measurement of instream nitrate concentration. Additionally, sampling equipment that automatically collects grab samples and conducts analyses that are traditionally done in the laboratory are increasing in availability and popularity (WET Labs, 2006; YSI, 2006; Jordan et al., 2007). Despite these developments, there are no current methods for the in situ, real time analysis of total phosphorus and total suspended solids for long term monitoring, so available data will remain spatially and temporally limited.

2.6 Surrogate measures

A surrogate is a measure that can be used to estimate another property such as contaminant concentration. A common surrogate measure in water bodies is turbidity, an optical measure of the scattering of light passing through a sample of water due to colloidal and suspended matter. Gray and Glysson (2002) report that in the United

States, turbidity is the most common measurement of water clarity and the most common surrogate of suspended sediment concentrations. Considerable research is available demonstrating the potential for accurately relating suspended sediment concentrations to turbidity measurements, some of which is subsequently described. In addition to using turbidity as a surrogate for suspended sediment, several studies have used in situ measurements as surrogates for other constituents that require laboratory analysis. As phosphorus is often closely associated with suspended solids, turbidity has been used as a surrogate for total phosphorus. Additional examples of in situ surrogates include turbidity as a surrogate for total nitrogen and fecal coliform and specific conductance as a surrogate for dissolved solids, alkalinity, sulfate, and chloride, as well as other ions.

Gray and Glysson (2002) asserted that suspended sediment loading and transport is more accurately calculated using the high frequency, continuous measurement of turbidity as a surrogate than by using sporadic measurements of concentration. Additional benefits of using surrogates to estimate suspended sediment loading include a decrease in the count of necessary grab samples, the potential for identifying sediment variability at a higher temporal resolution, and the ability to trigger automatic pumping samplers for the collection of samples for laboratory analysis (Gray and Glysson, 2002).

Through analyses of laboratory and field data, Gippel (1989, 1995) concludes that an acceptable correlation between field turbidity and suspended solids can generally be obtained, although the author warns that there can be confounding factors that influence the relationship. Specifically, turbidity is affected by the scattering properties of suspended particles, which are a function of particle size and composition. As a result,

the relationship between turbidity and suspended sediment will change with the source of sediment. Source material often varies from site to site and can fluctuate seasonally and even between storm events. Additionally, Gippel (1989) recommends the use of infrared turbidity sensors to eliminate the effects of water color on turbidity measurements.

Brasington and Richards (2000) used turbidity to monitor suspended sediment loads in five small catchments within the Likhu Khola basin in Nepal. The researchers examined both field and laboratory procedures to calibrate turbidity readings to suspended sediment concentrations, and then estimated sediment flux using both methods. The correlation using the field calibration was strong ($r^2 = 0.75$). One complication encountered by this study was the exceedence of the turbidity monitor's upper limit due to burial of the instrument by heavy sediment loads.

Christensen et al. (2000) used real time monitoring of turbidity, specific conductance, and discharge in conjunction with stepwise regression analyses to develop high frequency records of alkalinity, dissolved solids, total suspended solids, chloride, sulfate, atrazine, triazine, and fecal coliform for two sites on the Little Arkansas River in Kansas. A strong correlation between turbidity and total suspended solids was found at both sites throughout the four years of the study (correlation coefficients of 0.88-0.91). It was determined that two years of data consisting of 35 to 55 samples provided a sufficient sample size to correlate a constituent to its surrogate variables at these sites.

Uhrich and Bragg (2003) used turbidity as a surrogate for suspended sediment on three streams in northwestern Oregon. The relationship between turbidity and suspended sediment was strong at all sites (correlation coefficients of 0.90-0.93), and the

correlations were better than those observed between discharge and suspended sediment (correlation coefficients of 0.56-0.68).

Tomlinson and De Carlo (2003) used regression to relate continuous, high frequency turbidity with total suspended solids samples collected by automated samplers in three Hawaiian streams. Pooling all samples resulted in a correlation coefficient of 0.84, and separating the data by site improved the relationships (r^2 = 0.90-0.93). Several outlying points were attributed to non-uniformity between the sample collected and the water measured by the turbidity sensor. This could be because the automated sampler requires one full minute to collect a sample while the turbidity sensor measures instantaneously, or it could be due to distance between the sampler and sensor along the stream.

Grayson et al. (1996) conducted a study including data from multiple sites on the Latrobe River in Australia to determine if turbidity could reliably be used as a surrogate measurement for total phosphorus (TP) and total suspended sediments (TSS). The correlations developed were linear, and the correlation coefficients were 0.86 and 0.90 for TSS and TP, respectively. As the correlations were not site specific, error was introduced by spatial variability. The authors recognized that the preferred approach is to develop a relationship for each site that could be checked and adjusted after the initial period of sampling. The study concludes that turbidity is a more accurate estimator of TSS and TP than discharge alone and recommends that it be included in routine monitoring.

Kronvang et al. (1997) collected measurements of turbidity with TSS and particulate phosphorus over two years at two stations on Gelbaek Stream in Denmark. A strong correlation ($r^2 = 0.70$) between turbidity and TSS was observed during storm events, and the authors suggest that the correlation could be improved by developing seasonal relationships. Additionally, particulate phosphorus measured by this study was strongly correlated with suspended sediment concentrations ($r^2 = 0.87$) independent of seasonal variations.

Christensen (2001) used in situ measurements to estimate alkalinity, dissolved solids, TSS, suspended sediment, sodium, chloride, fluoride, sulfate, nitrate, total organic nitrogen, TP, and fecal coliform at one site on Rattlesnake Creek in Kansas. For both suspended sediment concentration and TSS, turbidity was the only explanatory variable, and the correlation coefficients were 0.825 and 0.926 respectively. For TP, the important surrogates were turbidity, specific conductance, and water temperature, and the correlation coefficient of the regression was 0.96. The authors suggest that water temperature provides a representation of season, and that the TP and turbidity relationship might vary between seasons. Specific conductance may be representative of changes in discharge, indicating that TP can depend on discharge as well as turbidity.

In a related study, Christensen et al. (2002) measured conductivity, pH, water temperature, dissolved oxygen, turbidity, and total chlorophyll using in situ monitoring equipment at four sites on three different Kansas rivers. Statistical regression was used to correlate the surrogates, or explanatory variables, to manually collected and measured concentrations of total nitrogen, TP, and fecal coliform. The authors point out that the

relationships developed for each constituent are site specific and may include different explanatory variables, although turbidity was common to all relationships. For TP, turbidity was the only important surrogate for three of the four stations while the regression at one station also included specific conductance and water temperature. Again, water temperature is probably important as it relates to seasonal variations in TP and specific conductance is likely related to discharge. Correlation coefficients for TP ranged from 0.51 to 0.96. Additionally, the authors tried to explain the relationship between an explanatory variable and the estimated constituent based on the hydrological characteristics and land use above each station.

Ryberg (2006) measured conductivity, water temperature, pH, turbidity, and dissolved oxygen in situ at one station on the Red River of the North in North Dakota. Manual water quality samples were collected and analyzed for alkalinity, dissolved solids, sulfate, chloride, nitrate/nitrite, total nitrogen, TP, and suspended sediment over three years. The study found that for suspended sediment, the important explanatory variables were turbidity and discharge, and the resulting correlation coefficient was 0.873. For TP, turbidity, discharge, and day of the year were the significant surrogates returning a correlation coefficient of 0.771. Ryberg suggests that the relationships are not consistent throughout seasons and that distinct relationships should be developed on a seasonal basis.

Stubblefield et al. (2007) examined turbidity as a surrogate for TSS, TP, and soluble reactive phosphorus at four locations on two low turbidity streams (0-50 NTU) in the Lake Tahoe Basin. Correlations were strong for the TSS and turbidity relationships

(correlation coefficients of 0.95 and 0.91), and not as strong, but still significant between TP and turbidity (correlation coefficients of 0.62 and 0.83). There was no significant correlation between turbidity and soluble reactive phosphorus. Due to the overestimation of TSS loads by discharge rating curves examined in this study, the authors determined that turbidity is a more accurate surrogate of TSS than is discharge. This is consistent with the findings of Phillips et al. (1999), who report that suspended sediment is subject to limitations in supply that are not reflected in the variability of discharge. Other studies have shown discharge to be an unsatisfactory surrogate for TP, as it can be affected by processes that are independent of hydrology such as biological uptake and incorporation into bottom sediments (Robertson and Roerish, 1999; Quilbe et al., 2006; Johnes, 2007; Jordan et al., 2007).

2.7 Sampling frequency

As mentioned, constraints on resources and logistics limit the frequency of grab sampling, and resulting concentration measurements are typically made at a frequency too low to accurately characterize constituent behavior that can change at time scales of less than one day (Kronvang and Bruhn, 1996; Horowitz, 2003; Tomlinson and De Carlo, 2003; Coynel et al., 2004). A number of studies have examined the effect of sampling frequency on load calculations, which are subsequently described. Richards and Holloway (1987), Kronvang and Bruhn (1996), Phillips et al. (1999), and Coynel et al. (2004) affirm that results from all methods of load estimation improve as sampling frequency increases. Several authors recommend continuous, high frequency monitoring in order to overcome uncertainty in load calculation resulting from infrequent sampling

and biased estimation methods (Ferguson, 1987; de Vries and Klavers, 1994; Johnes, 2007).

Richards and Holloway (1987) combined data to simulate a year of frequently collected data of TSS, TP, soluble reactive phosphorus, nitrate, and specific conductance. These data were subsampled to achieve sampling frequencies of four times per day, daily, weekly, semi-weekly, and monthly from which loads were calculated using two different equations for load estimation. Stratified sampling with additional samples collected during high flow periods was also examined. The correlation of the estimated load with the theoretical true load improved dramatically ($r^2 > 0.9$) with a sampling frequency of at least daily or with heavily stratified sampling regimes. The authors conclude that load estimates improve with increased sampling frequency, and that calculation method, watershed characteristics, and constituent behavior, as well as interactions between these factors, have a considerable effect on the results.

de Vries and Klavers (1994) found that sampling frequency was more important than load calculation equation for ammonium, chloride, and suspended matter for two different Dutch rivers using simulated time series with frequencies of 6, 12, 24, 52, 100, and 200 samples per year. For the smaller of the two rivers, none of the load estimation equations were deemed acceptable, even at the highest sampling frequencies (errors of ±25 percent). The authors recommend the investigation of alternative monitoring strategies such as automated samplers or in situ surrogate measures to achieve better load estimates.

Kronvang and Bruhn (1996) identify that there is no clear best load estimation method and that guidance is needed on sampling frequency and strategy. The authors used records of frequently (4 hour to one week intervals) sampled concentrations of nitrogen and phosphorus, which were subsampled to develop series of varying sampling frequencies. Reference loads were determined by linearly interpolating the concentration and discharge data to one minute to one day intervals. The subsampled time series were compared to the reference loads, and were also used to evaluate 13 load estimation equations. The authors conclude that the best method overall is a simple interpolation equation; however the most appropriate method depends on sampling frequency, constituent, and catchment. Although error was reduced as sampling frequency increased, the authors recommend fortnightly sampling in order to provide the greatest reduction in error while not being cost and resource prohibitive.

Phillips et al. (1999) examined the accuracy and precision of 22 load estimation procedures by applying each to weekly, fortnightly, and monthly datasets that were created by decimating 15-min records of discharge and suspended sediment as estimated from turbidity. The authors found significant variance between the results of each sampling method: for one site, the median value for weekly sampling varied between 21.3 and 105 percent of the reference load. The results indicate that for all sampling methods, precision consistently improves with increased sampling frequency. Additionally, no distinct estimation method was found to provide the most accurate results for all sampling locations and sampling frequencies. The authors conclude that

sampling at intervals of one week or greater does not provide sufficient accuracy or precision, regardless of the equation used to estimate the load.

Robertson and Roerish (1999) used annual loads calculated by interpolation as the reference to which they compared annual loads of TSS and TP estimated by discharge rating curves at sampling frequencies ranging from semi-monthly to sampling every six weeks. Increased sampling during periods of high flow was also examined. Using three different measures of error, the smallest bias was approximately 30 percent, and the range of the errors was greater than the typical interannual variability in the loads. This study concluded that for loads calculated by regression with discharge, a stratified sampling approach adds to bias rather than providing a better representation of constituent behavior at high flows. The goal of this research was to find a method for load calculation for streams where samples cannot be collected frequently, but a discharge rating curve method was deemed unacceptable. This study also found that load estimates depend not only on sampling frequency but on the length of the study and the hydrological conditions during the period of study.

Webb et al. (2000) developed seasonal synthetic concentration records using discharge rating curves for various constituents and watersheds. These records were then subsampled according to the frequency at which grab samples were actually collected, and loads were calculated by nine estimation equations. The equation which demonstrated the least amount of bias for each watershed and constituent was selected on a case by case basis to calculate resulting loads from which larger scale loads and yields were determined.

Coynel et al. (2004) simulated a range of sampling frequencies (four hour to one month intervals) of TSS and discharge based on a reference record of frequently (two hour to daily intervals) sampled TSS data. The authors found that the range of flux estimates decreased significantly as sampling frequency increased and reported that the error in monthly sampling can exceed the interannual variability in loads for extreme hydrological conditions. Using a threshold of ±20 percent, the authors determined that sampling at frequencies less than semi-weekly was unacceptable and that the required sampling frequencies were 7 hours for one watershed and 3 days for another.

Johnes (2007) used paired measurements of discharge and phosphorus at daily intervals in a variety of catchments as reference time series from which weekly and monthly records were derived. Eight equations for load estimation were applied to each subset of data, and the results were compared to the reference loads calculated from the reference series. For weekly sampling, the best method could only provide a "fuzzy estimate of TP load," and monthly sampling returned results biased by 50-450 percent, depending on the watershed and the calculation method. The author advocates further analysis using data at higher frequencies than daily.

2.8 Synopsis and objectives

Because TP is a constituent of concern on the Little Bear River, this study is examining high frequency surrogate measures as methods to estimate TP concentrations and loads with more certainty. High frequency monitoring with in situ sensors captures important periods in constituent transport and reveals short term variability as well as diurnal and seasonal trends typically omitted by intermittent sampling (Grayson et al.,

1997; Kronvang et al., 1997; Christensen et al., 2002; Tomlinson and De Carlo, 2003; Kirchner et al., 2004). Additional advantages to high frequency monitoring include overall reductions in costs, personnel, and logistics, increased spatial scales, and the potential to automate data collection (Grayson et al., 1997; Vivoni and Camilli, 2003; Kirchner et al., 2004; Pressl et al., 2004; Vivoni and Richards, 2005). As technology to measure TP in situ for extended time periods has not been developed, in situ turbidity can be used as a surrogate measure for TP.

Phosphorus is often associated with suspended solids, which are commonly estimated using in situ turbidity (Gray and Glysson, 2002). Several studies have also developed relationships between TP and turbidity (Grayson et al., 1996; Kronvang et al., 1997; Christensen, 2001; Christensen et al., 2002; Ryberg, 2006; Stubblefield et al., 2007). Although these studies have been conducted in watersheds of differing characteristics exhibiting a range of turbidity values, the relationships are site specific and are limited to a handful of streams. Furthermore, the authors advocate the development and implementation of surrogate relationships as a component of regular water quality monitoring programs.

High frequency sampling also eliminates that need to employ a complex equation to estimate loads. A number of studies have examined the various averaging methods used to calculate loads from infrequently sampled data (Richards and Holloway, 1987; de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Phillips et al., 1999; Robertson and Roerish, 1999; Webb et al., 2000; Coynel et al., 2004; Johnes, 2007). Many authors affirm the frequency of sampling is more important than the estimation method, and

recommend increased sampling frequency to avoid uncertainty introduced by calculation method, watershed characteristics, water quality constituent, and interaction between these factors (Ferguson, 1987; de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Phillips et al., 1999; Coynel et al., 2004; Johnes, 2007). Load estimates for larger rivers were generally found to be less biased than those of smaller rivers, for which increased sampling frequency is more important (Ferguson, 1987; Richards and Holloway, 1987; de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Phillips et al., 1999; Coynel, 2004; Johnes, 2007). The effect of timing of sample collection on load estimates was not explicitly addressed by any of these studies.

This study uses data from two sites in the Little Bear River to examine whether high frequency measures can be used to better understand constituent transport, what high frequency concentrations can reveal about the timing, sources, and pathways of TP and TSS transport, and whether the frequency and the timing of sample collection have an impact on load calculations. Based on the literature reviewed and described, the specific objectives of this research are:

1. Examine turbidity as a potential surrogate for TP and TSS.
2. Develop site specific equations to describe the relationship between turbidity and TP and TSS.
3. Examine other potential explanatory variables for significance in describing TP and TSS.
4. Generate high frequency estimates of TP and TSS concentrations using the relationships and high frequency turbidity data.

5. Using the high frequency TP and TSS concentrations along with associated discharge, calculate reference loads on an annual basis.
6. Subsample the high frequency concentrations and discharges to represent decreasing sampling frequencies and the timing of sample collection, calculate annual loads, and compare to the reference loads.

CHAPTER 3

SURROGATE MEASURES FOR PROVIDING HIGH FREQUENCY ESTIMATES OF TOTAL SUSPENDED SOLIDS AND TOTAL PHOSPHORUS CONCENTRATIONS[1]

Abstract

Surrogate measures, like turbidity, which can be measured with high frequency in situ, have potential for generating high frequency estimates of total suspended solids (TSS) and total phosphorus (TP) concentrations. In the Little Bear River, a semi-arid, snowmelt driven, and irrigation regulated watershed in northern Utah, USA, high frequency, in situ water quality measurements (turbidity, water level, and water temperature) were recorded in conjunction with periodic chemistry sampling conducted over a range of hydrologic conditions. Site-specific relationships were developed using turbidity as a surrogate for TP and TSS at two monitoring locations. Methods are presented for employing censored data in the regressions and for investigating explanatory variables in addition to the surrogate variables such as discharge conditions and storm events. Turbidity was a significant explanatory variable for TP and TSS at both the upper and lower watershed sites, which are characteristically different and have varying sources of discharge as well as phosphorus. At both sites, the relationships between TP and turbidity varied between spring runoff and baseflow conditions while the relationships between TSS and turbidity were consistent across hydrological conditions.

[1] Coauthored by Amber Spackman Jones, David K. Stevens, Jeffrey S. Horsburgh, and Nancy O. Mesner.

The methods developed in this paper enable the calculation of continuous, high frequency time series of TP and TSS concentrations that have previously been unavailable using traditional monitoring approaches. These methods have broad application for situations that require accurate characterization of the fluxes of these constituents over a range of hydrologic conditions.

3.1 Introduction

Traditional water quality monitoring programs rely on the analysis of grab samples that are typically collected at a frequency too low to fully characterize water quality constituent concentrations and to calculate loads of those constituents over time (Etchells et al., 2005; Scholefield et al., 2005). Additionally, concentrations of solids and nutrients are often greater during storm events due to non-point source runoff (Nolan et al., 1995; Kronvang et al., 1997, Correll et al., 1999; Croke and Jakeman, 2001; Houser et al., 2006; Jordan et al., 2007), periods that routine sampling often misses. High frequency monitoring with in situ sensors offers a number of enhancements to traditional water quality monitoring methods. High frequency monitoring can capture time periods and characterize seasonal trends that may be omitted or overlooked by traditional periodic grab sampling (Grayson et al., 1997; Christensen, 2001; Christensen et al., 2002; Tomlinson and De Carlo, 2003; Kirchner et al., 2004; Jordan et al., 2007). Monitoring equipment that measures continuously can reduce the logistics and personnel required for grab sampling to be representative (Grayson et al., 1997), can eliminate errors in transcription and delays in obtaining data (Vivoni and Camilli, 2003), and can be closely linked with a water quality model to better refine parameters and results (Vivoni and

Richards, 2005). Variables commonly measured in situ include physical parameters such as water level, pH, specific conductance, dissolved oxygen, and turbidity. Additionally, UV-VIS spectroscopy and ion-specific sensors can be used in situ to quantify constituents such as nitrate, nitrite, chlorophyll, and chemical oxygen demand.

Despite developments in sensor technology, there are still important water quality constituents that are either impossible or impractical to measure in situ or in real time for extended periods (e.g., total phosphorus samples are most often digested and analyzed in the lab). High frequency measurements have the powerful potential to be used as surrogates to estimate other properties such as pollutant concentrations. A common surrogate used for this purpose is turbidity, which is an optical measure of the scattering of light passing through a sample of water due to colloidal and suspended matter. This paper examines turbidity as a surrogate measure for total phosphorus (TP) and total suspended solids (TSS) at two locations on the Little Bear River, Utah, USA. We use the linear relationships between turbidity and TSS and TP to obtain equations for TP and TSS concentrations as functions of turbidity, enabling the generation of high frequency, long term estimates of their concentration.

Phosphorus is an essential nutrient in aquatic systems. However, over-enrichment of water bodies with phosphorus can cause increased primary productivity leading to eutrophication in lakes and reservoirs and excessive periphyton growth in rivers (Hem, 1985; US EPA, 1986; Mueller and Helsel, 1996). Concerns with eutrophic water bodies include aesthetics for natural waters and drinking water sources and reduced dissolved oxygen levels, which adversely affect fish and other forms of aquatic life. Phosphorus is

found naturally in some soils, but significant amounts are contributed to aquatic systems by anthropogenic sources such as fertilized fields, animal waste, wastewater treatment plants, and industries (Hem, 1985; Mueller and Helsel, 1996). Depending on the source, phosphorus is frequently associated with suspended sediments, which may also be a water quality concern (Kronvang et al., 1997; Heimlich, 2003). Not only do suspended sediments transport contaminants such as nutrients, pesticides, and metals, high levels of suspended sediment can be detrimental to aquatic life, decrease the recreational quality of a water body, complicate water treatment, and interfere with the operation of hydraulic structures (US EPA, 1986).

Considerable research is available demonstrating the potential for accurately relating suspended sediment concentrations to turbidity measurements (Gippel 1989, 1995; Kronvang et al., 1997; Brasington and Richards, 2000; Uhrich and Bragg, 2003; Christensen et al., 2000; Christensen, 2001; Lewis, 2002; Tomlinson and De Carlo, 2003). There is also evidence that turbidity can be used as a surrogate for phosphorus. Grayson et al. (1996), Christensen (2001), Christensen et al. (2002), Ryberg (2006), and Stubblefield et al. (2007) found statistically significant correlations between turbidity and TP in watersheds of differing characteristics exhibiting a range of turbidity values. In these studies, turbidity was the principle explanatory variable for TP and TSS, although the relationships at a few locations included discharge and a temporal variable (e.g., day of the year) in the final surrogate relationship. As the nature of turbidity depends greatly on the source of sediment (Gippel, 1995), the surrogate relationships are generally site

specific (Grayson et al., 1996; Christensen et al., 2002; Tomlinson and De Carlo, 2003), which limits the applicability of previous studies to other locations.

Surrogate relationships for estimating water quality constituent concentrations such as those presented in this paper allow for the generation of concentration estimates at a much higher temporal resolution than most traditional water quality monitoring programs have achieved. Although many aspects of water quality monitoring have improved, sampling frequency remains a limiting factor in the estimation of water quality constituent loads (de Vries and Klavers, 1994; Johnes, 2007). High frequency estimates of concentration can overcome some problems encountered when constituent loads are calculated (e.g., complicated load estimation equations and situations where discharge is measured more frequently than concentration). Water quality models also suffer from the paucity of concentration observations and would be improved by high frequency estimates of concentration (Neilson and Chapra, 2003; Kirchner et al., 2004; Johnes, 2007). As a result, compliance with water quality standards and regulations that are based on concentration and load estimates can be determined with more certainty.

Surrogate measures can be an important component of water quality monitoring programs and environmental observatory design as a relatively inexpensive method for producing high frequency time series of water quality constituent concentrations over extended time periods. The Little Bear River is one of 11 environmental observatory test bed projects developing techniques and technologies for environmental observatory design ranging from innovative application of environmental sensors to publishing observations data in common formats and making it widely accessible (Montgomery et

al., 2007). Specific objectives of the Little Bear River Test Bed include the estimation of water quality fluxes from surrogate data, relation of the fluxes to watershed attributes and management practices, examination of high frequency hydrologic and hydrochemical responses, and development of cyberinfrastructure supporting these analyses.

In this paper, we describe the development of surrogate relationships for TP and TSS at two locations in the Little Bear River. Section 2 describes the Little Bear River watershed where this study was conducted. Section 3 details the data collection and statistical procedures used to obtain the surrogate relationships. Section 4 includes the final surrogate models and a comparison of the two sites.

3.2 Study area

The Little Bear River watershed is located in northern Utah, USA and is a major tributary of the Bear River, which flows into the Great Salt Lake. The Little Bear watershed encompasses an area of approximately 740 km^2, the headwaters are in the Bear River Mountain Range, and elevations range from 1,340 m to 2,700 m. The river has two principal subdrainages, the East Fork and the South Fork. The South Fork and its major tributary, Davenport Creek, flow northward through forest and range land before the confluence with the East Fork. The East Fork originates in higher elevation, forested land, and flows northwest until it is contained by Porcupine Reservoir, which is used to store water for summer agricultural irrigation. A few miles downstream of Porcupine dam, the East Fork is diverted for irrigation purposes, and for several months of the year, portions of the natural channel are dry. The confluence of the two forks is near the town of Avon, after which the river flows northward through the towns of Paradise and Hyrum.

Most of the land adjacent to the river is agricultural including crops and livestock grazing. Near the town of Hyrum, the river is contained in Hyrum Reservoir, which is also operated to supply summer irrigation water. Below Hyrum dam, the river flows northwest through lower gradient agricultural land. The river passes through the towns of Wellsville and Mendon before draining into an arm of Cutler Reservoir and ultimately to the Bear River. The watershed and local towns are shown in Fig. 3-1.

Over the past 15 years (1993-2007), the average annual precipitation in the lower watershed was 432 mm, while the average annual precipitation in the upper watershed was 4,465 mm, demonstrating significant variability in annual precipitation with elevation. Most of the precipitation occurs as snowfall, and the flow regime in the watershed is driven by snowmelt with hydrograph peaks occurring in late spring. The magnitude, timing, and duration of the peak are dictated by the winter snowpack and spring weather conditions. In the upper watershed, where an active United States Geological Survey (USGS) gage is located, the average annual discharge is 2.5 cms (based on 15 years of data), and within a year, the discharge ranges from 0.50 to 12 cms on average.

3.3 Methods

3.3.1 Instrumentation and monitoring

Seven sites have been instrumented within the Little Bear River for the collection of high frequency water quality monitoring data. General characteristics and data collected at these locations are described in detail by Horsburgh et al. (2008). Two of these sites were chosen for analysis in this paper and are indicated in Fig. 3-1. The first

site is the Little Bear River at Paradise, located in the upper watershed below the confluence of the East and South Forks and above Hyrum Reservoir. The second site is the Little Bear River at Mendon, located in the lower watershed near the river's terminus at Cutler Reservoir. The two sites were selected for their distinct characteristics. Above Paradise, there are agricultural diversions and the river passes through some agricultural land, but relative to Mendon, the river is less regulated, higher gradient, and less impacted. In contrast, above Mendon, the river is controlled by Hyrum reservoir releases and influenced by agricultural return flows, a wastewater treatment plant, and an increasingly agriculturally developed landscape. Approximately 4 percent of the land above Paradise is agricultural whereas between Paradise and Mendon, the portion of land used for agriculture is about 50 percent. Additionally, at Mendon, the river is lower gradient and groundwater levels in this portion of the watershed are higher than at Paradise. Another difference between the two sites is characteristics of the soils and resulting suspended sediments. Mendon is located in a lacustrine valley with finer soils that remain in suspension while the suspended matter at Paradise is coarser and more likely to settle (Soil Survey Staff, 2008).

The water quality monitoring equipment installed at both sites includes a Forest Technology Systems DTS-12 SDI-12 Turbidity Sensor. The turbidity sensor uses an infrared light beam and optical backscatter with a detector at 90 degrees to the emitted light to determine turbidity (Forest Technology Systems Ltd., 2007), and the sensor also measures water temperature. Turbidity and water temperature measurements were recorded at half hour intervals. At Paradise, there is an active USGS gage (USGS

10105900 Little Bear River at Paradise, UT) adjacent to the real time water quality sensors from which records of 15-minute instantaneous and daily average discharge were obtained. At Mendon, water level is measured continuously by a KWK Technologies SPXD-600 SDI-12 Pressure Transducer. Water level readings were coupled with periodic manual discharge measurements to obtain a stage-discharge relationship. The stage-discharge relationship was then used to generate continuous, half hourly estimates of discharge at Mendon.

Water quality samples were collected at the two sites either by grab sampling conducted by a field crew or by automated samplers. The samplers operate by pumping water from the river through tubing into sample bottles held within the main chamber, allowing for the collection of multiple samples during an event such as a storm or a period of snowmelt. In general, samplers were deployed when precipitation was expected. Each sample was split for TSS and TP analysis with a portion of the sample filtered using a 0.45 μm filter for the analysis of dissolved total phosphorus (DTP).

Laboratory analyses were performed externally by labs affiliated with Utah State University and with the State of Utah Division of Water Quality. This study uses historic data, so labs and their associated methods changed over the time period examined. The results from the labs should produce consistent results, and a small number of samples sent to multiple labs confirmed this assumption. For TSS analyses, some samples were analyzed under EPA method 340.2, Total Suspended Solids, Mass Balance while the remaining samples were analyzed according to EPA method 160.2, Residue Nonfilterable Total Suspended Solids. For TP and DTP analyses, some samples were analyzed

according to EPA method 200.8, Determination of Trace Elements in Water and Waste by Inductively Coupled Mass Spectroscopy, and the remaining samples were analyzed as directed by EPA method 365.2, Orthophosphate Ascorbic Acid Manual Single Reagent preceded by an acid digestion of the sample.

3.3.2 Database procedures

All of the mentioned datasets were stored and managed using a database at the Utah Water Research Laboratory (http://littlebearriver.usu.edu/). The turbidity, water temperature, and water level data were transmitted and accessed via a spread spectrum radio network, the USGS discharge data were obtained from the USGS National Water Information System (NWIS) and incorporated into the database, and the lab results were entered into the database by hand. The time period under examination extended from the installation of in situ sensors in August 2005 through April 2008 resulting in datasets of 150-180 samples of TP, DTP, and TSS collected at each site. For each observation of TP, DTP, and TSS, associated continuous measurements were extracted from the database and matched in time with the lab results. When the timing of a sample did not exactly correspond to the timing of continuous measurements, the values of turbidity, water temperature, and discharge that bracketed the manual sample were interpolated accordingly.

3.3.3 Statistical methods

Our objective was to develop correlations to estimate TP and TSS as functions of turbidity using simple regression, following the general form given in Equation 3.1.

$$y_i = \alpha_0 + \alpha_1 x_i + e_i \quad i = 1,2, \ldots n \tag{3.1}$$

where y_i represents the ith observation of the response variable, α_0 and α_1 are parameters estimated by regression, x_i is the ith observation of the explanatory variable, e_i represents the error for the ith observation, and n is the number of samples. Using techniques subsequently described, regression parameters unique to each response variable were estimated based on the observations datasets. The errors, or residuals, should be independent, demonstrate constant variance, have a mean of zero, and be normally distributed. Examining the residuals for these qualities helps in assessing the appropriateness of the developed equation.

In order to assess the potential of turbidity as a surrogate for TP and TSS, we initially examined plots of turbidity against the response variables. This allowed for the visual identification and subsequent removal of several extreme data points (no more than 3.5 percent of a single dataset). The outliers consisted of high turbidity measurements corresponding to low TP or TSS measurements, as well as low turbidity measurements corresponding to high TP or TSS measurements, relative to the majority of data points. It is assumed that these outliers are a consequence of inconsistency between grab samples and the water that passes in the range of the turbidity sensor. Although efforts were made to collect samples near the turbidity sensors, there could still be discrepancy between the collected sample and the water measured by the turbidity sensor. This is consistent with the findings of Christensen et al. (2000) and Tomlinson and De Carlo (2003).

While turbidity was thought to be a significant explanatory variable, other variables (discharge, water temperature, day of year, and hour of day) were considered

for inclusion in the regression equations and were tested for significance in describing some of the variability in the response variables. In addition to these parameters, categorical variables associated with the hydrological conditions at the time of sample collection were examined.

Categorical variables are qualitative descriptors of the data that can be used in regression models. For each data point, a categorical variable is assigned a value of 1 or 0 to designate whether or not the observation falls into a particular category (e.g., seasons, laboratory methods). By adding categories as explanatory variables to the regression equation, this technique permits the inclusion of multiple categories while developing a single model that describes the entire dataset. The alternative of splitting the data into subsets according to categories and developing multiple models is less statistically powerful as resolution is lost with a reduced number of observations (Berthouex and Brown, 2002). Equation 3.2 shows the form of a regression equation with the inclusion of a categorical variable.

$$y_i = \alpha_0 + \alpha_1 x_i + Z(\beta_0 + \beta_1 x_i) + e_i \quad i = 1,2, \ldots n \tag{3.2}$$

where Z represents a categorical variable ($Z = 0$ if data are in the first category, $Z = 1$ if data are in the second category), β_0 and β_1 are parameters estimated by regression, and y_i, α_0, α_1, x_i, e_i and n are as defined previously.

In this study, we investigated two categorical variables associated with hydrological conditions: one to represent spring runoff versus baseflow and one to represent the occurrence of a storm. Because the flow regime of the Little Bear is primarily snowmelt driven, we hypothesized that the behavior of TP and TSS might be

significantly different during spring runoff versus baseflow conditions. Seasonal differences in surrogate relationships have been suggested by Grayson et al. (1996), Christensen et al. (2002), and Ryberg (2006). As a result, the first categorical variable that was examined was whether the sample was collected during baseflow conditions or during spring runoff. Observations identified to occur during the period of spring runoff were assigned a value of 1 for this variable while the remaining observations, collected during baseflow conditions, were assigned a value of 0. Since runoff resulting from precipitation events also has the potential to carry significant amounts of sediment and associated phosphorus into the river, the other categorical variable that was hypothesized to be significant was whether or not a sample was collected during a storm event. Observations identified as occurring during a storm event were assigned a value of 1 for this variable, and all other observations, collected during non-storm periods, were assigned a value of 0.

Initially, storms were defined as any time appreciable precipitation occurred based on a record of daily precipitation in the lower watershed. However, even though efforts were made to sample during precipitation events, rainfall often occurred without a significant discharge response, especially during the summer months when antecedent soil conditions were not conducive to runoff generation. As a result, alternative methods were employed to determine whether a sample was collected during a storm event. One technique examined was baseflow separation. Baseflow separation refers to the partitioning of a hydrograph into baseflow (i.e., discharge due to groundwater sources) and runoff (i.e., discharge that is a response to an external event such as a storm or

snowmelt). There are many methods for baseflow separation ranging from simple to complex (McCuen, 1998; Chapman, 1999). We needed to identify runoff that was a response to precipitation and not a result of spring snowmelt, so the local-minimum method was selected because it delineates more of the discharge as baseflow relative to other baseflow separation techniques, making it more appropriate for separating storms from other sources of discharge. This method was used as part of a publicly available program known as the Web-based Hydrograph Analysis Tool (WHAT) to perform baseflow separation (Lim et al., 2005).

Storm identification is somewhat subjective, and because we wanted to explore simple methods, additional techniques were developed to designate samples collected during a storm. One method was based on a visual examination of the discharge and precipitation records. Another method was based on a reference distribution for averages of sets of consecutive observations, which can reveal the significance of change in a serially correlated data series (Berthouex and Brown, 2002). Table 3.1 details the methods used for storm identification. Storms were identified separately for each site, and both high frequency and daily average discharge data were employed to assess whether higher resolution provides a superior method of identification or if lower resolution better represents the period of the river's response to a precipitation event. Each sample was assigned a value (0 or 1) for storm category for each method, and the resulting datasets were tested for significance as an explanatory variable in the regression equation for each response variable.

A significant portion of the TP concentrations were reported as non-detects (30 percent at Paradise and 13 percent at Mendon), so we needed a regression method with the capability to include censored data. Historically, censored data have either been omitted from analyses or substituted with some value at or below the detection limit. These methods introduce bias and variability into descriptive statistics that are calculated from datasets with censored values (Helsel, 2005). In order to preserve the censored values in the dataset without using substitution, regression with maximum likelihood estimation (MLE) was performed on the matched datasets within the framework of the statistical program R (http://www.r-project.org/) using techniques developed and described by Helsel and Lee (2006). MLE assumes a distribution for the response variable and estimates a mean and standard deviation for that dataset that are most likely to result in the values above the detection limit and the proportion of values below the detection limit (Helsel, 2005). The mean and standard deviation are then used to produce values for the regression parameters (e.g., α_0, α_1) that account for censored data. The TSS datasets do not suffer from a large amount of censored data, so associated models were developed using standard least squares regression within the R framework.

In order to determine which variables were important predictors, regression was performed multiple times for each response variable by adding and removing potential explanatory variables. A number of techniques were employed to address the appropriateness of each resulting model and to compare one model to another. For each explanatory variable in the regression, a p-value was calculated, indicating the probability that the value of the regression parameter is not different from 0, so a p-value greater than

a specified threshold (commonly 0.05) indicates that the relationship between the explanatory variable and response variable is not statistically significant. If a p-value was less than 0.05, the associated variable was considered significant. For MLE regression, overall log-likelihood tests assist in determining whether the model is better than no model at all, and a parallel test, the partial log-likelihood, was used to discern whether the addition of a variable improved the regression as compared to the equation without that variable (Helsel, 2005). The partial log-likelihood was then compared to a chi-square distribution with the associated degrees of freedom to determine the p-value, the probability that the model with the additional variable was different than without it. Again, 0.05 was used as the criteria for significance. Finally, residuals were examined to assess the error in each model. The root mean square error (RMSE) as given by Equation 3.3 was used to compare models as a lower RMSE indicates a reduction in overall error.

$$RMSE = \sqrt{\frac{\sum r^2}{\upsilon}} \tag{3.3}$$

where *RMSE* is the root mean square error, *r* represents each residual value, and υ corresponds to the degrees of freedom. A variable was included in the final equation if it provided a significant reduction in the RMSE, had a p-value less than 0.05, and was significant according to the partial log-likelihood test. Plots of the residuals were also examined to verify randomness and independence from other factors as well as to assess whether the residuals exhibited constant variance and approached a normal distribution.

Transformations are often used on datasets to achieve constant variance, a linear relationship between independent and dependent variables, or a normal distribution in the residuals (Berthouex and Brown, 2002). A log transformation of the dependent variable

alone as well as a log transformation of both the dependent and independent variables were examined. Transformations did not provide any significant improvement in the models for any of the response variables, so untransformed datasets were used in all cases.

3.4 Results and discussion

3.4.1 Simple correlations

Plots of the relationships between TP and TSS and potential explanatory variables are shown as matrices of correlation plots in Fig. 3.2, 3.3, 3.4, and 3.5. At Paradise, there is a strong correlation between turbidity and both TP and TSS (correlation coefficients of 0.95). Both response variables exhibit some correlation with discharge (TP correlation coefficient of 0.80 and TSS correlation coefficient of 0.70) and water temperature (TP correlation coefficient of 0.48 and TSS correlation coefficient of 0.30). Additionally, there appears to be some relationship with the day of the year on which the sample was collected for both TP (correlation coefficient of 0.57) and TSS (correlation coefficient of 0.46).

At Mendon, TP appears to have a significant correlation with turbidity (correlation coefficient of 0.70), though not as strong as that at Paradise nor as strong as the correlation between turbidity and TSS at Mendon (correlation coefficient of 0.84). Like Paradise, TP and TSS at Mendon have some correlation with day of year (TP correlation coefficient of 0.67 and TSS correlation coefficient of 0.56). In addition, TSS at Mendon is correlated with discharge (correlation coefficient of 0.41).

3.4.2 Paradise: Total phosphorus

The final model for TP at Paradise is given by Equation 3.4:

$$TP = 0.0209 + 0.000798 * Turb + 0.0386 * Z \quad (3.4)$$

where *TP* is total phosphorus concentration in mg/L, *Turb* is turbidity in NTU, and *Z* represents the categorical variable for spring runoff (*Z* = 1) versus baseflow (*Z* = 0). The p-value for turbidity was less than 10^{-6} and for *Z* was $8.71*10^{-4}$, both within the 0.05 threshold. Excluding the residuals of the censored data, the RMSE for this model was 0.069 mg/L TP, which is about a fourth of the MLE mean of the observed dataset, 0.26 mg/L TP. This value is within the range of RMSE values resulting from the turbidity and TP correlations reported by Christensen et al. (2002) over a similar range of turbidity values.

Of all the explanatory variables examined, only turbidity and the spring runoff categorical variable were significant. Including discharge or water temperature did not improve the equation's ability to predict TP concentrations. This is likely due to colinearity with turbidity, as shown in Fig. 3.2. The relationships between TP and discharge and water temperature are very similar to the relationships between turbidity and discharge and water temperature. The categorical variable indicating whether observations were collected during a storm event was not significant regardless of which storm identification method was used. This implies that the relationship between turbidity and TP is consistent throughout storm events, though there is a distinction during periods of spring runoff and periods of baseflow. Correlation with season or

hydrologic regime is consistent with the results of Christensen et al. (2002) and Ryberg (2006).

Fig. 3.6 contains plots of observed TP and modeled TP using Equation 3.4. Fig. 3.6(a) shows a time series of modeled TP along with points of observed TP for the entire period used to generate Equation 3.4. However, this plot does not permit direct comparison between each point as does Fig. 3.6(b), which indicates corresponding modeled and observed results connected by vertical lines. Fig. 3.6(b) shows that the differences between the regression results and the observations are generally greater at higher values of TP, but there are exceptions to this pattern.

Fig. 3.7 and 3.8 show plots of the residuals of this model. Fig. 3.9 is a matrix of correlation plots showing relationships between the residuals and measured physical/chemical properties. These plots and the associated correlation coefficients do not demonstrate a strong relationship between the residuals and any measured properties. Additionally, the residuals did not show any correlation with temporal variables such as day of the year or hour of the day. In the interest of space, residual plots for the other equations in this paper are found in Appendix A.

Although the other criteria for residuals are met (independence, constant variance, and mean of zero), the probability plot indicates that the residuals of this model are not normally distributed. Non-normal residuals suggest that the assumed parametric distribution (in this case, a normal distribution) in the regression is incorrect. However, normality in the residuals was not achieved through logarithmic transformations of TP or turbidity, and the un-transformed model returned the smallest RMSE. The incorporation

of any additional explanatory variables did not provide a significant increase in the normality of the residuals. Because of the non-normality of the residuals, techniques that assume normality (e.g., confidence limits on the slope) cannot be conducted on this data. In order to verify that using MLE parametric regression was valid, Kendall's tau and the associated Akritas-Thiel-Sen (ATS) line were calculated. These methods can be used to non-parametrically determine the slope between an independent and dependent variable (Helsel and Lee, 2006). For TP at Paradise, the ATS slope corroborated the coefficient between turbidity and TP determined by MLE regression.

3.4.3 Paradise: Total suspended solids

The final model for TSS at Paradise is given by Equation 3.5:

$$TSS = 3.58 + 1.31 * Turb \tag{3.5}$$

where *TSS* is total suspended solids in mg/L and *Turb* is turbidity in NTU. The p-value for turbidity was less than 10^{-6}, within the criteria for significance. Turbidity was the only explanatory variable that was a significant descriptor of TSS, suggesting that turbidity alone is sufficient to predict TSS across hydrologic conditions at this site. As with TP, although there appears to be a correlation between TSS and discharge and water temperature, the correlation between turbidity and these variables is similar (see Fig. 3.3), so the relationship with turbidity provides an adequate estimate.

The resulting RMSE was 117 mg/L TSS, which is about half of the dataset mean of 240 mg/L TSS. Plots of the modeled and the observed datasets are found in Fig. 3.6. Plots of the residuals, a histogram of the residuals, the residual probability plot, and a matrix of correlation plots of the residuals and physical/chemical variables are found in

Appendix A (Fig. A.1, A.2, and A.3). Fig. 3.6(d) shows results similar to those of TP at Paradise with greater errors at higher values of TSS. Like the TP model at Paradise, the distribution of the residuals deviates from the normal at the tails, but the ATS slope verified the regression parameters. Significant correlation is not observed between the residuals and any additional variables.

3.4.4 Mendon: Total phosphorus

The final model for TP at Mendon is given in Equation 3.6:

$$TP = -0.0341 + 0.0053 * Turb + 0.0949 * Z - 0.00404 * Turb * Z + 0.0832 * Y - 0.00871 * Y * Turb \quad (3.6)$$

where TP is total phosphorus concentration in mg/L, *Turb* is turbidity in NTU, Z represents the categorical variable for spring runoff ($Z = 1$) versus baseflow ($Z = 0$), and Y is a categorical variable for *Turb* <10 ($Y = 1$) versus *Turb*>10 ($Y = 0$). The p-values for turbidity, Z, and the interaction between turbidity and Z were all less than 10^{-6} and all within the criteria for significance. This equation differs from that for the Paradise site in that the interaction between turbidity and Z was found to improve the model significantly, indicating that the combined effect of the two variables is different from the sum of their individual contributions. In this case, TP is decreased during spring runoff periods by a factor of 0.00404*turbidity. This reduction, however, resulted in some negative predicted concentrations, so an additional categorical variable, Y, was included to distinguish the relationship at low versus high levels of turbidity. The p-value for Y was $1.38*10^{-3}$, and the p-value for the interaction between Y and turbidity was $5.24*10^{-3}$, both within the 0.05 threshold. The inclusion of this variable suggests that the relationship between turbidity and TP is different at low values of turbidity, corresponding to low TP

measurements. Distinctions in surrogate relationships at low turbidity levels have been suggested by Grayson et al. (1996) and Stubblefield et al. (2007).

As with TP at Paradise, turbidity and the spring runoff/baseflow categorical variable were the only explanatory variables that were found to be significant. In Fig. 3.4, TP at Mendon shows little correlation with water temperature or discharge, although there is correspondence with day of the year, which is corroborated by the inclusion of *Z* in the final equation. Similar to Paradise, none of the storm event variables resulted in improvement in the model, demonstrating that at this site, the relationship between turbidity and TP varies between the spring runoff and baseflow periods, but is consistent through storm events.

Without using the residuals of the censored values, the RMSE for this model was 0.027 mg/L TP, which is about a third of the MLE mean of the observed values (0.074 mg/L TP). This value is less than the range of RMSE values determined for turbidity and TP relationships by Christensen et al. (2002) for a similar range of turbidity values. Fig. 3.10 includes plots of the modeled results with the observations, which, unlike the results at Paradise show a wide range of errors at the low and high values of TP. Residual plots are found in Appendix A (Fig. A.4, A.5, and A.6). The residuals appear to be independent from other variables, and the probability plot indicates that the distribution of the residuals closely approximates the normal distribution.

3.4.5 Mendon: Total suspended solids

The final model for TSS at Mendon is given by Equation 3.7:

$$TSS = 0.341 + 1.41 * Turb \tag{3.7}$$

where *TSS* is total suspended solids in mg/L and *Turb* is turbidity in NTU. The p-value for turbidity was less than 10^{-6}, which surpasses the criteria for significance. In parallel with the TSS model at Paradise, the Mendon model was not significantly improved by the inclusion of any explanatory variables other than turbidity. Although there was some correlation between TSS and discharge at this site, it is reflected in the relationship between turbidity and discharge, so no new information is gained by including discharge in the equation. Additionally, neither of the categorical variables was significant, implying that the relationship between TSS and turbidity is consistent through hydrological conditions at this site.

The RMSE of the model was 10.8 mg/L TSS, which is about a third of the mean of observed values (30.4 mg/L TSS). Plots of the modeled and observed data are presented in Fig. 3.10. Like both TP and TSS at Paradise, Fig. 3.10(d) shows a greater difference in modeled and observed values at higher values of TSS, though there are exceptions to this generalization. Plots of the residuals are found in Appendix A (Fig. A.7, A.8, and A.9). The residuals demonstrate randomness in relation to explored independent variables, and the probability plot shows that the residuals closely approximate the normal distribution although there is some deviation at the tails.

3.4.6 Site Comparison

Paradise and Mendon were selected as sampling sites for analyses in this paper due to their differing characteristics, which are reflected somewhat in the variations in the surrogate relationships. The final equations are summarized in Table 3.2. The RMSEs for both TP and TSS are greater at Paradise than Mendon, a result of the larger range of

observed values at that site. The simple correlations indicate stronger correlations at Paradise than Mendon between turbidity and both TP and TSS, and the final TP equation at Paradise appears to better track trends through a greater range than does the Mendon model. Also, the Mendon TP regression is more complex as it includes the interaction between turbidity and the spring runoff/baseflow categorical variable and requires an additional variable to account for different behavior between turbidity and TP at low concentrations.

We hypothesize that these differences are a result of the varying composition of TP between the two sites. Of the TP measurements with corresponding DTP measurements, at Mendon, an overall average of 60 percent of the TP was dissolved, leaving 40 percent as particulate. The average composition of TP measured at Paradise was 40 percent dissolved and 60 percent particulate, the opposite of the ratio at Mendon. These ratios are comparable to those reported by Johnes (2007) for sites with higher baseflow and more groundwater influence (65-75 percent dissolved) versus those with lower baseflow (40-50 percent dissolved). Since dissolved phosphorus is not associated with any particles, the correlation between TP and turbidity at Mendon would not be as strong as the correlation at Paradise where the TP is primarily comprised of particulate phosphorus. This is corroborated by Stubblefield et al. (2007) who found no correlation between soluble reactive phosphorus and turbidity.

It is inferred that the variations in the speciation of phosphorus at the two sites is a reflection of the different sources of phosphorus and differing stream dynamics. Factors that may increase the amount of DTP at Mendon include more concentrated agricultural

activity than above Paradise, impact from a wastewater treatment plant, and manure or fertilizer that flows into canals and into the river before being incorporated by plants or adsorbed to the soil. In contrast, it is hypothesized that the phosphorus entering the river above Paradise is primarily related to soil erosion and particulate matter. Additionally, between the two sites is Hyrum Reservoir. Phosphorus (primarily particulate) enters the reservoir from the upper watershed and accumulates in the lake bed. Over time, the phosphorus can dissolve and then be carried out of the reservoir in its dissolved form through reservoir releases (Utah DEQ, 2000b). Reservoir releases might also carry algae that contain phosphorus. Other than releases from Hyrum Reservoir, the sources of discharge at Mendon include agricultural return flows, which have the potential to contribute dissolved phosphorus from crop runoff, and there is some groundwater influence at Mendon as well. It is possible that dissolved phosphorus enters the river via the groundwater (Burkart et al., 2004), however we have no specific evidence that this is occurring in the Little Bear River.

Patterns of dissolved phosphorus were further examined in order to address the possibility of relating the portion of TP that was dissolved to model results. No trends were found with respect to season, and there was no relationship with TP model residuals at either site. During runoff periods at both sites, the fraction of dissolved phosphorus was slightly higher than during baseflow periods, but more data is necessary to confirm that these differences are significant.

At Mendon, the TP surrogate relationship might be improved by the inclusion of variables in addition to turbidity, discharge, and water temperature. Part way into this

study, in situ sensors were installed to measure pH, specific conductance, and dissolved oxygen. Because specific conductance and pH are related to dissolved species, they may help to refine the regression where the majority of TP is dissolved. A method for using surrogate measures to estimate dissolved phosphorus would be valuable because, although TP is the form of phosphorus that is generally regulated, dissolved phosphorus is the form that is actually available for biological uptake.

Despite the differing characteristics of the two sites, aspects of the surrogate relationships were consistent between Paradise and Mendon. At both sites, the TSS surrogate relationships were functions only of turbidity with similar coefficients (1.31 at Paradise and 1.41 at Mendon) although the suspended matter differs between the two locations. At Paradise, the soils and resulting suspended solids are coarser and more likely to settle than the finer material that is more likely to stay suspended found at Mendon. For both sites, turbidity was the only explanatory variable for TSS while the TP relationship included a variable to account for baseflow versus spring runoff. This is similar to the differences between the suspended sediment and TP surrogate relationships determined by Ryberg (2006). Another similarity is the lack of significance of storm event in all of the regressions. Although storms are often important periods for TSS and TP transport and despite significant investment into the identification of storm periods, in this watershed, the relationships between turbidity and TP and TSS do not vary during storms.

3.5 Conclusions

Physical (turbidity, discharge, water temperature) and temporal (day of year, hour of day) variables were matched with TP and TSS observations at two sites in the Little Bear River to assess the potential for using continuously measured turbidity as a surrogate for estimating TP and TSS. Regression equations were developed for TP and TSS as functions of turbidity at both sites. In developing the surrogate relationships, censored data were employed using MLE regression, and categorical variables representing hydrological conditions were investigated. We found that the relationships between turbidity and the response variables were not significantly improved by the use of a categorical variable indicating storm events versus no storm. At both locations, however, there was a distinction in the relationship between turbidity and TP during periods of spring runoff versus periods of baseflow. For TSS, the relationship with turbidity was consistent across hydrological conditions at both locations. At the lower watershed site, the TP model included a distinction between low and high levels of turbidity. The overall error in the models, as estimated by the RMSE, was between one fourth and one half of the mean of the observed data, and visual examinations of the observed and estimated concentrations indicate that the equations generally track observed trends.

The differences in the surrogate relationships at the upper and lower watershed sites allude to differences in the sources of phosphorus as well as sources of discharge. Turbidity was the only physical variable that was a significant surrogate, although, since the inception of this study, additional variables (pH, specific conductance, dissolved

oxygen) have been measured that may improve the relationships. Furthermore, as new data are collected, the regression equations may be modified, or calibrated, to improve the fit with observations.

Coupled with high frequency measurements of explanatory variables, surrogate relationships can be used to calculate high frequency estimates of concentration for extended time periods. Loads derived from high frequency, continuous concentration records provide a number of advantages to loads calculated from traditionally sampled concentration. One benefit is that increased loading during events such as storms or spring runoff, which are often missed by routine sampling programs, are considered without skewing the estimate high as collecting samples disproportionately during storm events can do. Also, there is no need to use complicated load estimation equations that allow for long periods between concentration measurements or discharge measured more frequently than concentration.

Surrogate measures to estimate water quality constituents have widespread implications for water quality monitoring programs and the design of environmental observatories. Until viable in situ sensors for TP and TSS are developed, surrogate measures allow the characterization of fluxes at varying time scales (e.g., seasonally or in response to an individual event) and also provide a better means for comparison between monitoring sites. High frequency concentration estimates and resulting loads will allow the determination of compliance based on a concentration or load threshold to be made with more certainty. For water quality models, improved quantification of constituent concentrations will facilitate the estimation of parameters representing pollutant loading

drivers such as land use, management practices, and hydrologic characteristics and can also permit the testing of underlying model assumptions. For large scale environmental observatories, the use of surrogate measures will be necessary as a logistically and economically feasible means to characterize the variability in constituent fluxes on high temporal and spatial resolutions over extended time periods and at many locations.

Table 3.1
Methods used for storm identification

Method	Description
Baseflow High Frequency	Baseflow separation was performed with the highest frequency discharge. If runoff was greater than 5%* of the baseflow for that time increment, the observation was identified as having occurred during a storm.
Baseflow Daily Average	Baseflow separation was performed with the daily average discharge. If runoff was greater than 5%* of the baseflow for that day, the observations were identified as having occurred during a storm.
Visual High Frequency	Samples were identified as having been collected during a storm where precipitation occurred and where there was a visibly notable change in the hydrograph of the highest frequency discharge data.
Visual Daily Average	Samples were identified as having been collected during a storm where precipitation occurred and where there was a visibly notable change in the hydrograph of daily average discharge data.
Reference Distribution	If the increase in three day average discharge was greater than 5%* of the change in average discharge of the previous three days, then observations on that day was identified as having occurred during a storm.

* Five percent was selected as a threshold to provide a reasonable amount of change from baseflow to represent a true response to a precipitation event.

Table 3.2
Final surrogate equations

Site	Constituent	Equation
Paradise	Total Phosphorus	$TP = 0.209 + 0.000798 * Turb + 0.0386 * Z$
	Total Suspended Solids	$TSS = 3.58 + 1.31 * Turb$
Mendon	Total Phosphorus	$TP = -0.0341 + 0.0053 * Turb + 0.0949 * Z - 0.00404 * Turb * Z + 0.0832 * Y - 0.00871 * Y * Turb$
	Total Suspended Solids	$TSS = 0.341 + 1.41 * Turb$

Variable	Description
TP	Total Phosphorus, mg/L
TSS	Total Suspended Solids, mg/L
$Turb$	Turbidity, NTU
Z	categorical variable for spring runoff ($Z = 1$) versus baseflow ($Z = 0$)
Y	categorical variable for $Turb$ <10 NTU ($Y = 1$) versus $Turb$ >10 NTU ($Y = 0$)

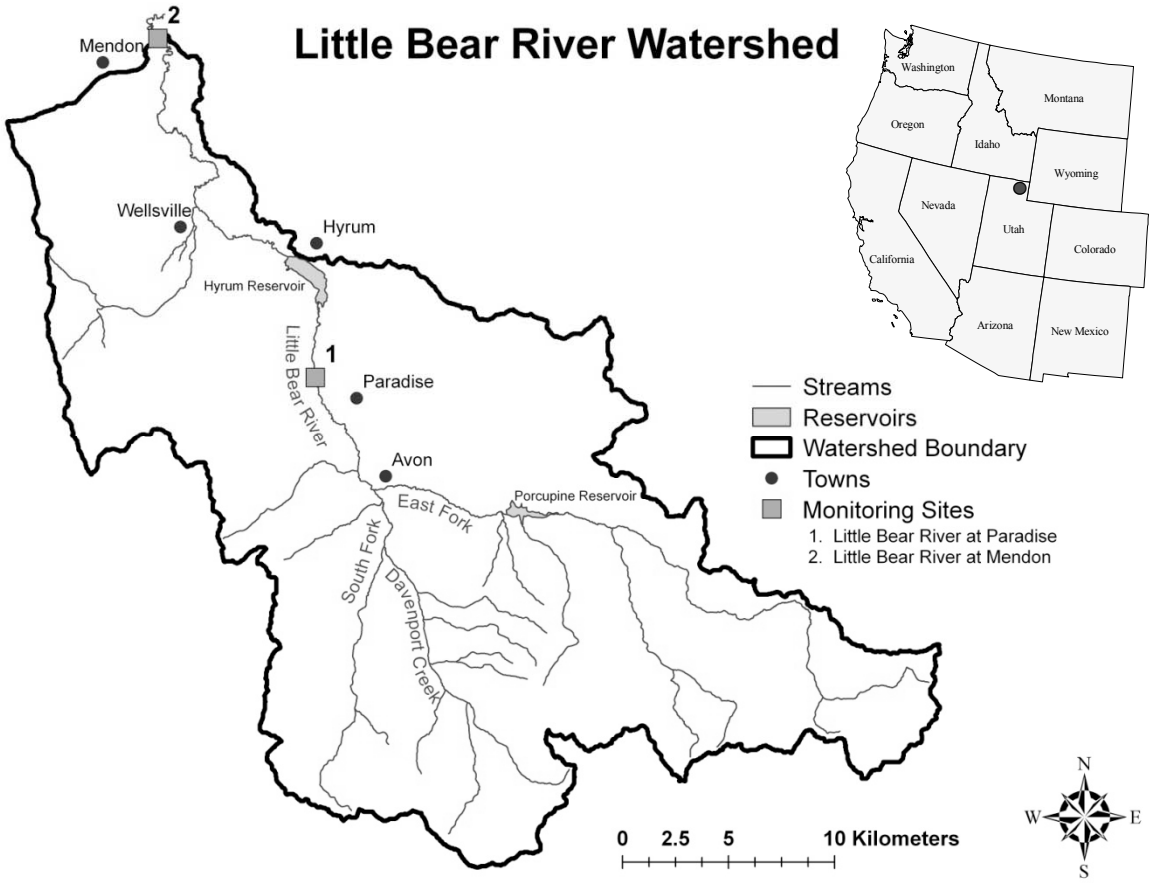

Fig. 3.1. Little Bear River watershed.

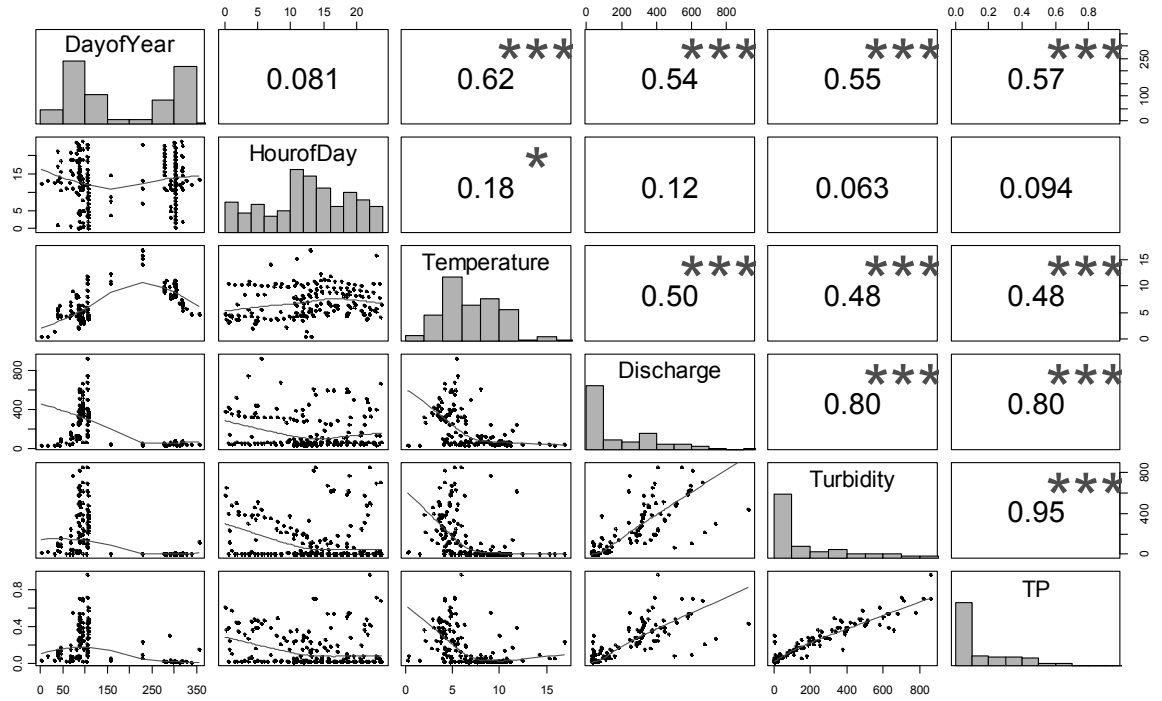

Fig. 3.2. Paradise TP correlation matrix. Stars indicate the significance of the Pearson's correlation coefficient (Three stars is significant to the 0.001 level, two stars is significant to the 0.01 level, and one star is significant to the 0.05 level.)

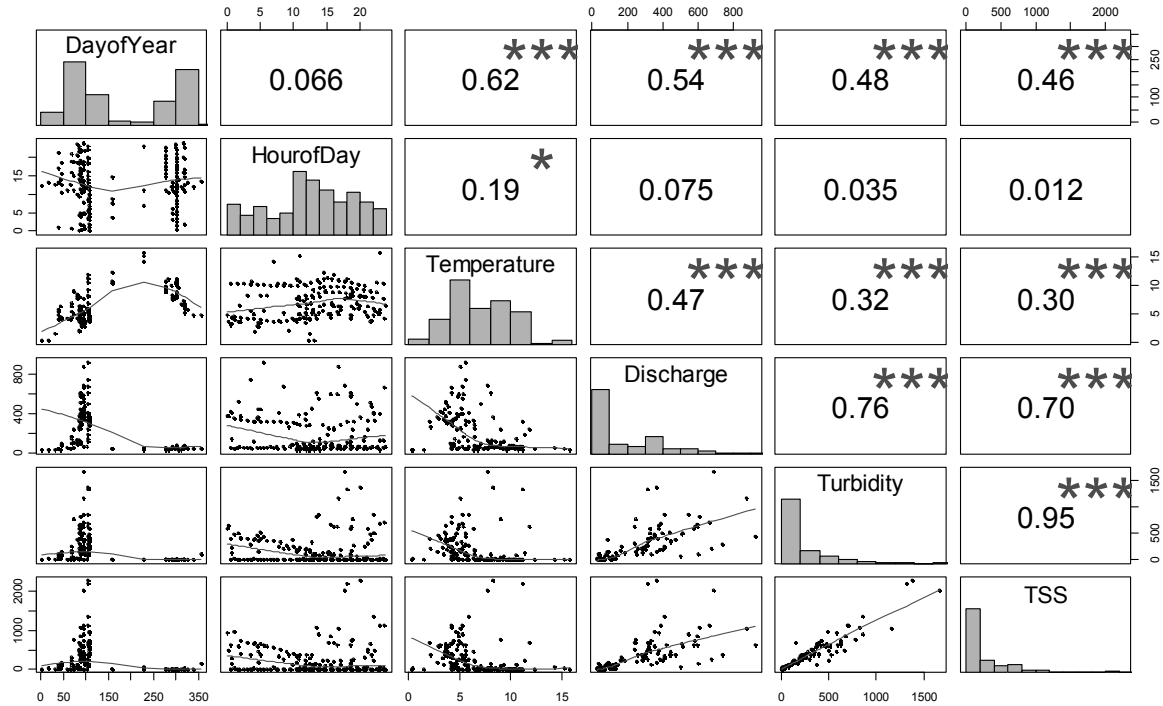

Fig. 3.3. Paradise TSS correlation matrix. See Fig. 3.2 for interpretation of stars.

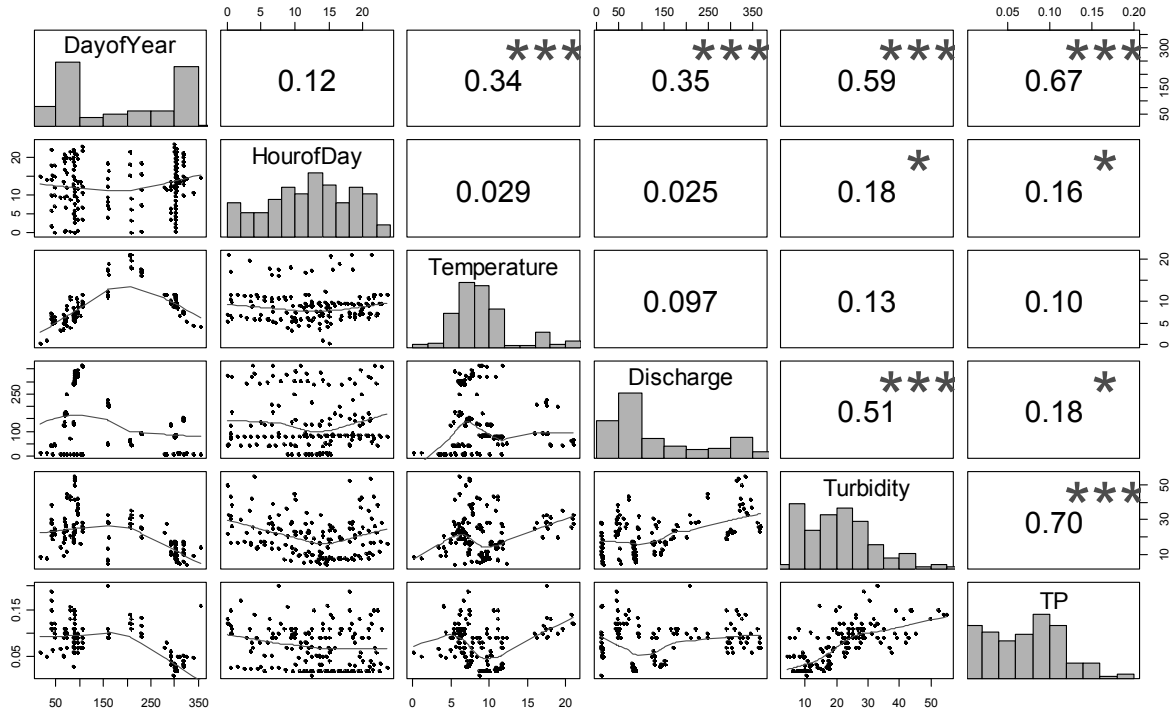

Fig. 3.4. Mendon TP correlation matrix. See Fig. 3.2 for interpretation of stars.

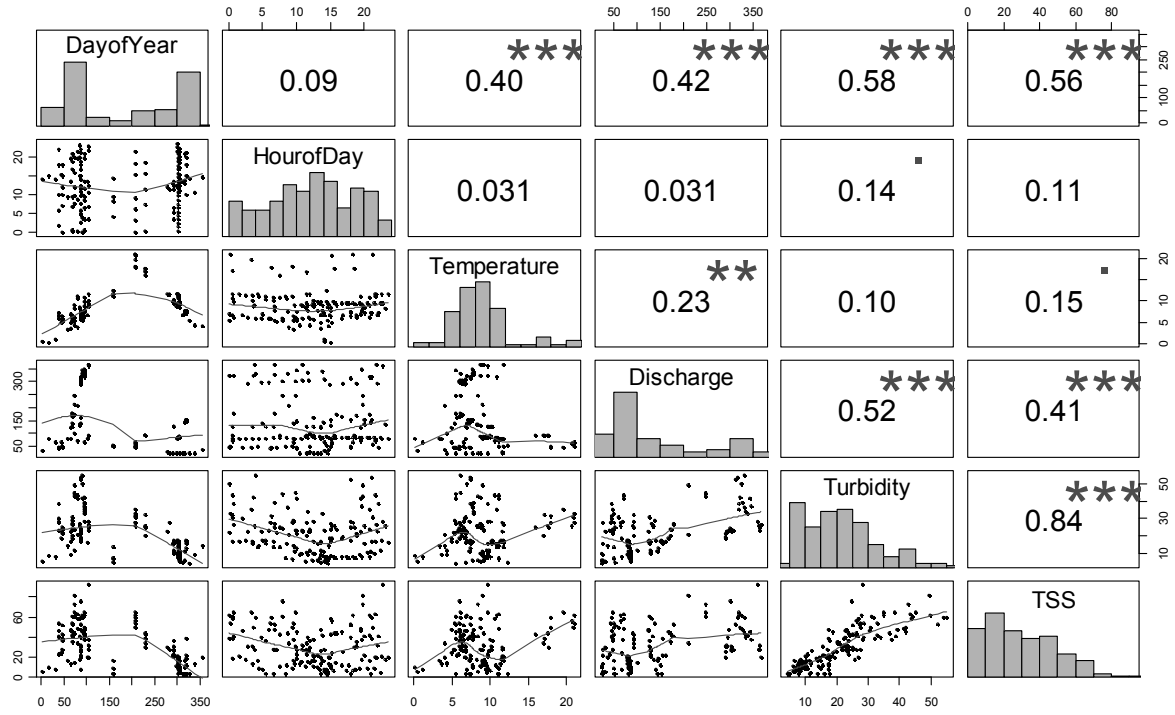

Fig. 3.5. Mendon TSS correlation matrix. See Fig. 3.2 for interpretation of stars.

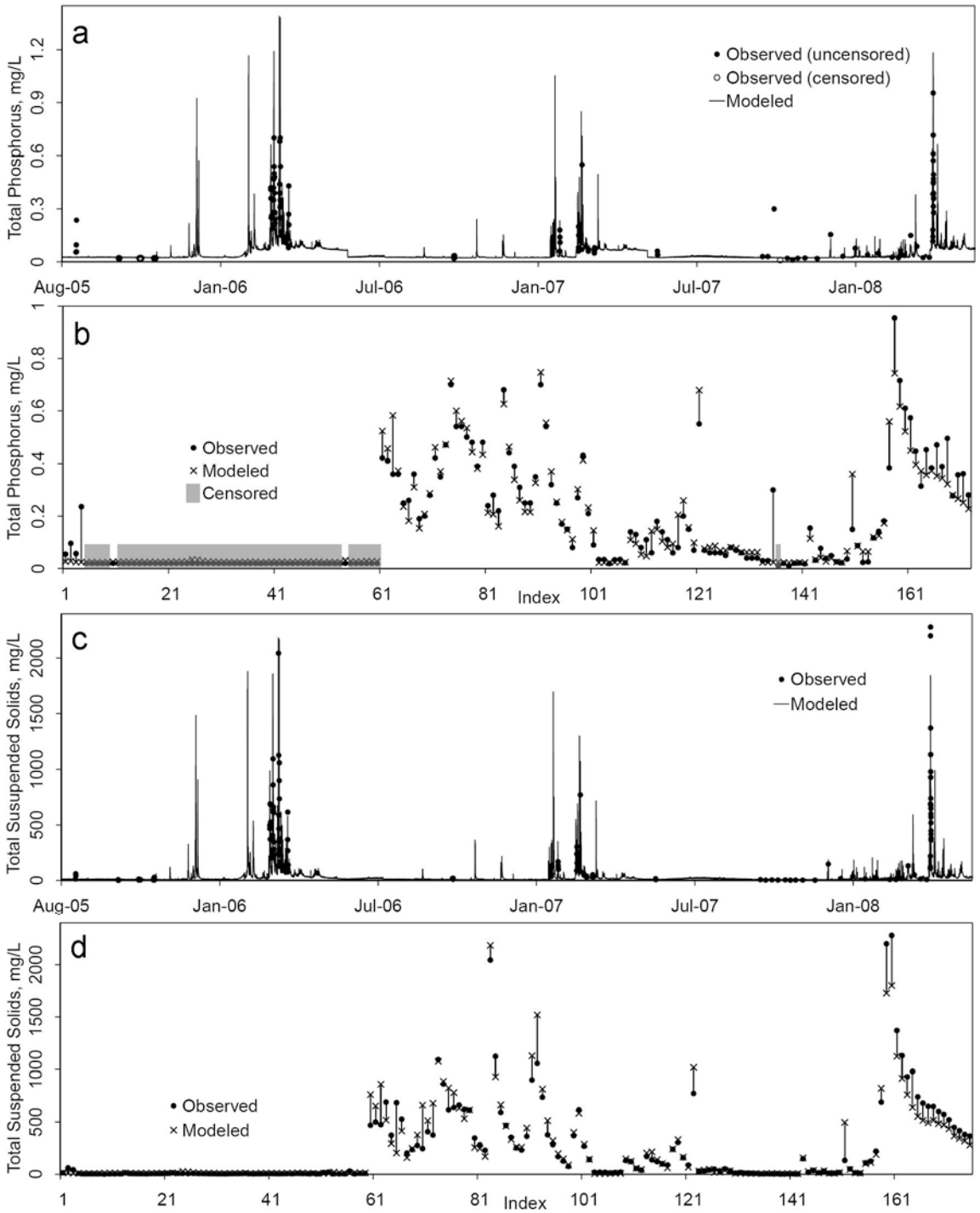

Fig. 3.6. Plot of observed and modeled TP (a and b) and TSS (c and d) at Paradise. For censored data, points are plotted at the detection limit. As many of the observation in the full time series (a and c) are obscured, panels b and d only contain modeled results with a corresponding observation. Observed and modeled values are connected by vertical lines. The x-axis is an index that represents the order in which observations were made.

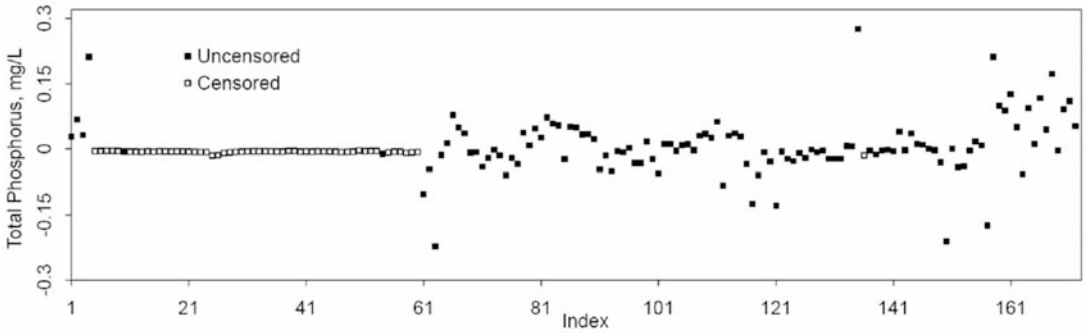

Fig. 3.7. Residuals of the Paradise TP model.

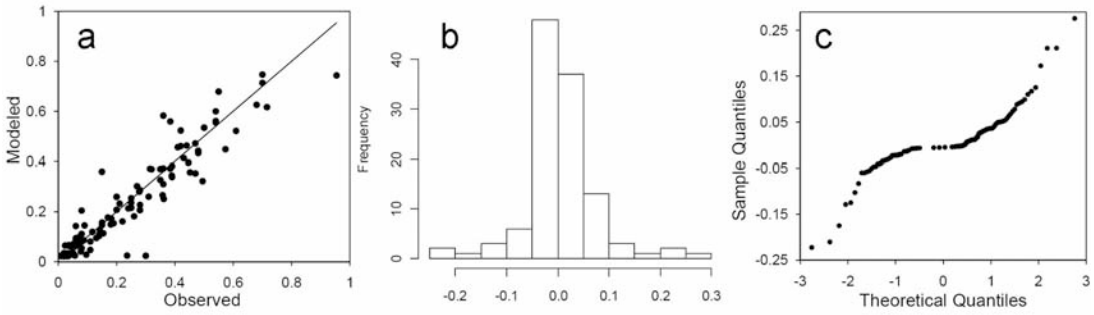

Fig. 3.8. Statistical plots for the TP model at Paradise. Observed versus modeled TP (a), a histogram (b) and a probability plot (c) of residuals.

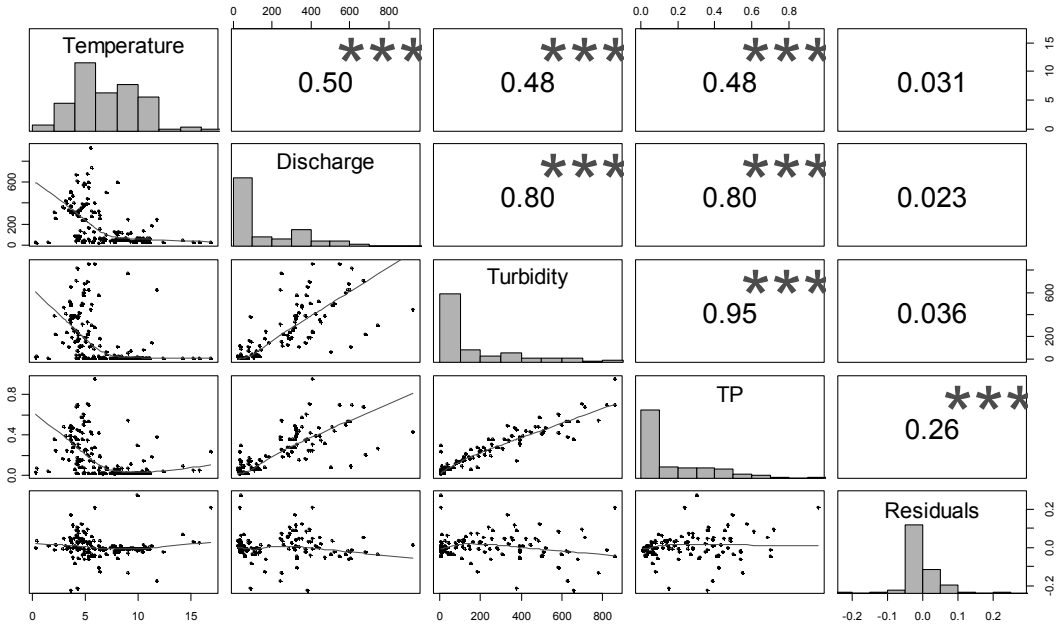

Fig. 3.9. Residuals of the Paradise TP model compared with measured variables. See Fig. 3.2 for interpretation of stars.

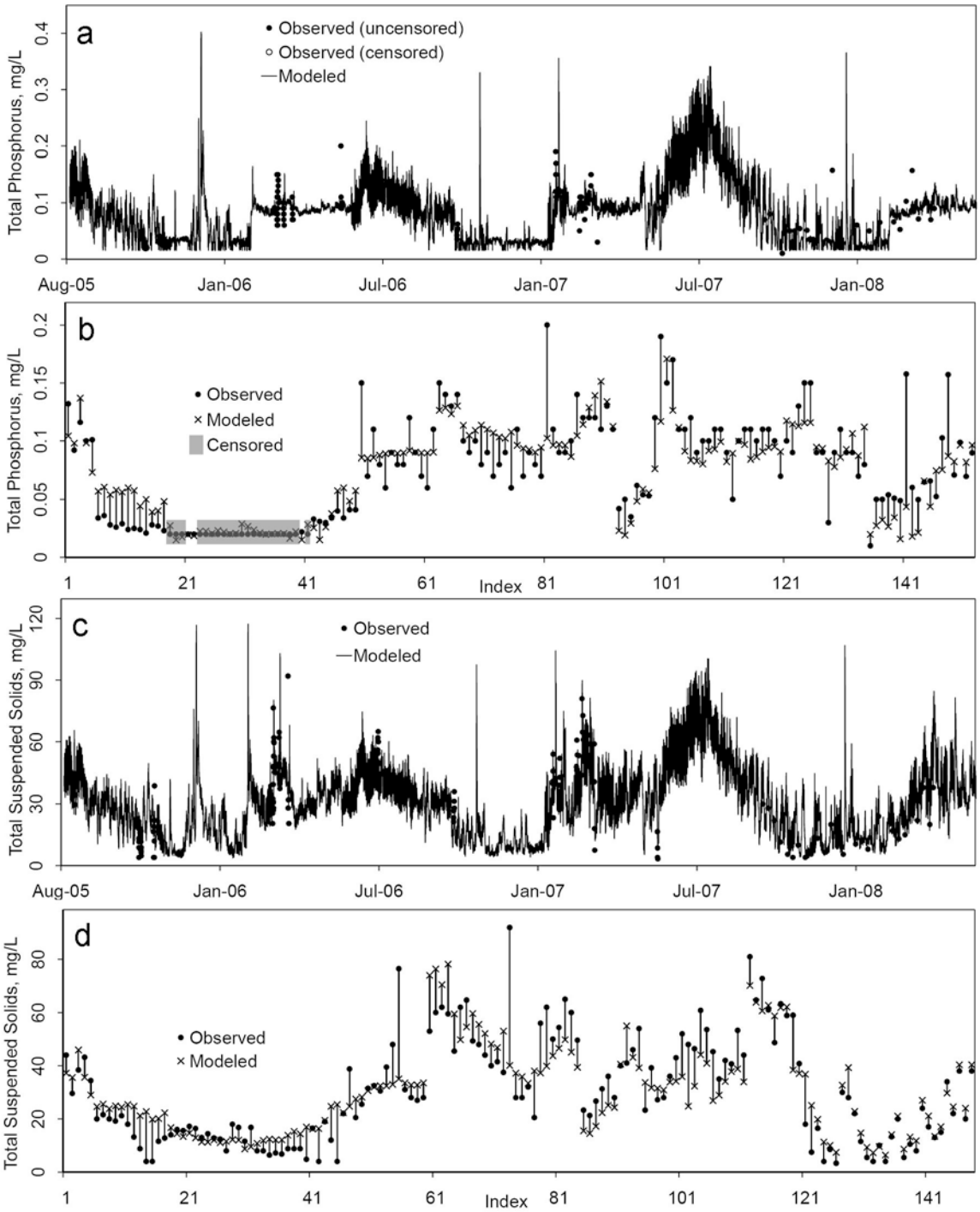

Fig. 3.10. Plot of observed and modeled TP (a and b) and TSS (c and d) at Mendon. For censored data, points are plotted at the detection limit. As many of the observation in the full time series (a and c) are obscured, panels b and d only contain modeled results with a corresponding observation. Observed and modeled values are connected by vertical lines. The x-axis is an index that represents the order in which observations were made.

CHAPTER 4

IMPACT OF SAMPLING FREQUENCY ON ANNUAL LOAD ESTIMATION OF TOTAL PHOSPHORUS AND TOTAL SUSPENDED SOLIDS[2]

Abstract

Compliance with water quality standards for sediment and nutrients is typically based on the collection and analysis of grab samples. These data generally are not collected with enough frequency or regularity to provide representation of the constituent loading, yet regulatory decisions and the investment of significant resources for water quality improvement are routinely based upon these numbers. In the Little Bear River watershed in northern Utah, USA, continuous, high frequency measurements of turbidity were used to generate high frequency estimates of instream total phosphorus (TP) and total suspended solids (TSS) concentrations through surrogate relationships. The concentration estimates were paired with discharge data to estimate TP and TSS loading (reference loads). The high frequency records were then subsampled to create random subsets representing hourly, daily, weekly, and monthly sampling frequencies. Additionally, subsets were created to examine the effects of randomizing the time of day and the day of week of sampling. The annual load estimates resulting from the decimated subsets were compared to the reference loads. Results show that high frequency surrogate measures generally improved estimates of TP and TSS loads in comparison to grab sampling. Overall, higher frequency sampling resulted in load estimates that better approximated the reference loads, although the amount of bias varied between sites.

[2] Coauthored by Amber Spackman Jones, Jeffrey S. Horsburgh, Nancy O. Mesner, Ronald J. Ryel, and David K. Stevens.

Additionally, the hour of the day and the day of the week on which sampling is conducted can have an impact on load estimation, depending on sampling location and hydrologic conditions.

4.1 Introduction

Water quality constituent loadings are generally determined through the collection and analysis of concentration grab samples along with instantaneous estimates of discharge. For most water quality monitoring programs, the frequency of grab sampling requires a balance between the necessary resolution to estimate accurate loads and the resource costs of increased sampling (Kronvang and Bruhn, 1996; Horowitz, 2003; Coynel et al., 2004). Furthermore, the frequency required for grab sampling to be representative of constituent behavior may be logistically infeasible due to the number of samples that will have to be collected and analyzed (Coynel et al., 2004). Water quality models require concentration observations for calibration and also suffer from sparse concentration datasets (Neilson and Chapra, 2003). Compliance with water quality regulations is often determined by grab sampled concentrations and resulting loads, even when important periods for constituent transport may be omitted (Jordan et al., 2007). Although many aspects of water quality monitoring and modeling have improved, sampling frequency is and is likely to remain a limiting factor in load estimation and water quality modeling (de Vries and Klavers, 1994; Kirchner et al., 2004; Johnes, 2007).

Various equations have been proposed for the calculation of loads given discrete measurements of concentration and discharge. de Vries and Klavers (1994), Kronvang and Bruhn (1996), Phillips et al. (1999), Etchells et al. (2005), and Johnes (2007)

compared results from various estimation equations and provide direction on equation selection. These studies, among others, conclude that the most appropriate equation for load calculation depends on watershed characteristics, hydrological behavior, the nature of the constituent, the frequency of sample collection, and interactions between these factors (Richards and Holloway, 1987; Kronvang and Bruhn, 1996; Robertson and Roerish, 1999; Webb et al., 2000; Johnes, 2007). Additionally, discharge is commonly measured at higher frequency than concentration, adding complexity to load calculations because concentration cannot be discretely paired with discharge (de Vries and Klavers, 1994; Robertson and Roerish, 1999; Kirchner et al., 2004).

No single equation for load estimation has been found to provide acceptably unbiased and precise results across a range of conditions (Kronvang and Bruhn, 1996; Phillips et al., 1999); furthermore, uncertainty can be introduced by virtue of the equation selected (Johnes, 2007). However, Richards and Holloway (1987), Kronvang and Bruhn (1996), Phillips et al. (1999), and Coynel et al. (2004) affirm that results from all equations improve as sampling frequency increases.

Several authors recommend high frequency, continuous monitoring in order to overcome uncertainty in load calculation resulting from infrequent sampling and biased estimation methods (Ferguson, 1987; de Vries and Klavers, 1994; Quilbe et al., 2006; Johnes, 2007). In addition to eliminating the need to select one of many complex load estimation equations, high frequency monitoring provides a number of advantages over traditional grab sampling. High frequency monitoring captures periods that are often overlooked by routine sampling and overcomes the logistic challenges required for

representative sampling (Grayson et al., 1997; Christensen, 2001; Christensen et al., 2002; Tomlinson and De Carlo, 2003; Kirchner et al., 2004; Jordan et al., 2007).

This paper examines the effect of sampling frequency on load calculations using random subsets of high frequency concentration estimates to simulate periodic grab sampling at different frequencies. For two sites in the Little Bear River in northern Utah, USA, regression relationships were developed using turbidity as an explanatory variable for total phosphorus (TP) and total suspended solids (TSS) that consider censored data as well as hydrological conditions (Spackman Jones et al., 2008). These relationships were used to construct continuous, high frequency (half hour interval) time series of estimated TP and TSS concentrations at two sites. In this paper, we describe the results of decimating the synthetic concentration records at varying intervals to create time series subsets that simulate periodic grab sampling. Each resulting subset was used to calculate associated annual loads for two years of data. These loads are compared to the reference loads calculated from the original synthetic concentration record. Section 2 describes the Little Bear River watershed and the sites at which loads were calculated. Section 3 describes the methods that were used in deriving the concentration time series, decimating the datasets, calculating loads, and evaluating the results. Section 4 relates the results of the load calculations for each scenario, compares them to the reference loads, and compares results between the two sites.

4.2 Study area

The Little Bear River watershed is located in northern Utah, USA and is a major tributary of the Bear River, which flows into the Great Salt Lake. The Little Bear watershed encompasses an area of approximately 740 km^2, with headwaters in the Bear River Mountain Range, and elevations in the watershed that range from 1,340 m to 2,700 m. The river has two principal subdrainages, the East Fork and the South Fork. The South Fork and its major tributary, Davenport Creek, flow northward through forest and range land before the confluence with the East Fork. The East Fork originates in higher elevation, forested land, and flows northwest until it is contained by Porcupine Reservoir, which is used to store water for summer agricultural irrigation. A few miles downstream of Porcupine dam, the East Fork is diverted for irrigation purposes, and for several months of the year, portions of the natural channel are dry. The confluence of the two forks is near the town of Avon, after which the river flows northward through the towns of Paradise and Hyrum. Most of the land adjacent to the river is agricultural including crops and livestock grazing. At Hyrum, the river is contained in Hyrum Reservoir, which is also operated to supply summer irrigation water. Below Hyrum dam, the river flows northwest through lower gradient agricultural land, passing through the towns of Wellsville and Mendon before draining into an arm of Cutler Reservoir. The watershed and local towns are shown in Fig. 4.1.

Over the past 15 years (1993-2007), the average annual precipitation in the lower watershed was 432 mm, while the average annual precipitation in the upper watershed was 4,465 mm, demonstrating significant variability in annual precipitation with

elevation. Most of the precipitation occurs as snowfall, and the flow regime in the watershed is driven by snowmelt with hydrograph peaks occurring in late spring. The magnitude, timing, and duration of the peak are dictated by the winter snowpack and spring weather conditions.

This paper examines loads calculated at two locations on the Little Bear River, which are indicated on Fig. 4.1. The first site is the Little Bear River at Paradise, which is located in the upper watershed below the confluence of the East and South Forks and above Hyrum Reservoir. The second site is the Little Bear River at Mendon, which is located in the lower watershed near the river's terminus at Cutler Reservoir. The two sites were selected for their distinct characteristics. Above Paradise, there are agricultural diversions, and the river passes through some agricultural land, but relative to Mendon, the river is less regulated, higher gradient, and less impacted by human activity. In contrast, above Mendon, the river is controlled by Hyrum reservoir releases and influenced by agricultural return flows, a wastewater treatment lagoon, and an increasingly agriculturally developed landscape. Approximately 4 percent of the land above Paradise is agricultural whereas between Paradise and Mendon, agriculture accounts for about 50 percent of total land use. Additionally, at Mendon, the river is lower gradient and groundwater levels in this portion of the watershed are higher than at Paradise. The characteristics of the soils and resulting suspended sediments also differ between the two sites. Mendon is located in a lacustrine valley with finer soils that remain in suspension whereas the suspended matter at Paradise is coarser and more likely to settle (Soil Survey Staff, 2008). Differences between the two sites are also evident in

discharge records. Mendon generally has higher baseflow discharge with attenuated peaks while the discharge at Paradise is flashier. For the two years that comprise the period of this study, the mean discharge at Paradise was 2.5 cms with a maximum of 29 cms, and at Mendon the average discharge was 3.5 cms with a maximum of 11 cms.

4.3 Methods

A number of studies have artificially decimated reference datasets in order to compare the effect of sampling frequency on load estimates. The reference datasets for TP and TSS in these studies are generally based on infrequently sampled data (Johnes, 2007), although some authors generated higher frequency data through interpolation (Kronvang and Bruhn, 1996) or discharge rating curves (Webb et al., 2000), which have been shown to be an unsatisfactory estimator of TP and TSS (Phillips et al., 1999; Robertson and Roerish, 1999; Quilbe et al., 2006; Johnes, 2007; Jordan et al., 2007). Stubblefield et al. (2007) found turbidity to be a more accurate surrogate for TSS than discharge. In this papter, the reference datasets are high frequency estimates of TP and TSS concentrations calculated from turbidity, which are subsampled to examine the effects of sampling frequency on load estimates. The timing of sampling, which few studies have addressed, is also investigated.

4.3.1 Discharge and concentration time series

For both sites, high frequency discharge records were matched in time with concentration estimates to calculate loads. At Paradise, an active United States Geological Survey gage (USGS 10105900 Little Bear River at Paradise, UT) measures

instantaneous discharge at 15-minute increments. There were a number of time periods with gaps in the data that were filled by interpolation (if the period was less than 48 hours) or by substituting the daily average discharge (as obtained from the USGS record) for all values on that day. As the concentration time series consist of values every 30 minutes, only the discharge observations on the hour and the half hour were used to calculate loads. The time series of discharge at Paradise is shown in Fig. 4.2(a).

At Mendon, water level is measured every half hour by a KWK Technologies SPXD-600 SDI-12 Pressure Transducer. The water level measurements were paired with manually measured discharges in order to develop a stage-discharge relationship, which was then used to calculate a half hourly time series of discharge estimates. There were a few periods of missing data at this site as well, though none of them exceeded 48 hours, and values were interpolated accordingly. The time series of discharge at Mendon is shown in Fig. 4.3(a).

Concentrations of TP and TSS were estimated using site specific relationships with turbidity. Intermittently sampled, laboratory analyzed concentrations of TP and TSS were matched with corresponding turbidity values. The turbidity was measured every half hour at each site using a Forest Technology Systems DTS-12 SDI-12 Turbidity Sensor. The statistical program R (http://www.r-project.org/) was used to perform regression analysis on these data, returning functions that use turbidity to estimate each response variable (TP and TSS) at each site. As a significant portion of the TP data were reported as non-detects, maximum likelihood regression, which accounts for censored data, was used to generate the TP relationships (Helsel and Lee, 2006). The final

regression equations are presented in Table 4.1, and more details can be found in Spackman Jones et al. (2008). A small number of gaps in the turbidity data were filled by interpolation. At one site, there was a period of approximately six weeks in the summer missing turbidity data due to probe malfunction. Because turbidity is low and relatively constant during the summer, this gap was filled with data from the same dates of an adjacent year. The resulting turbidity data series, shown in Fig. 4.2(b) and 4.3(b) were used as input in the equations in Table 4.1 to generate high frequency time series of concentration estimates. Time series of the concentrations of TP and TSS at Paradise are shown in Fig. 4.2 and plots of the concentrations at Mendon are shown in Fig. 4.3.

Fig. 4.2 and 4.3 help demonstrate the differences between the two sampling locations. The upper watershed site (Paradise) is more heavily influenced by snowmelt, as indicated by significant peaks in discharge, turbidity, and concentration in late spring. In contrast, at the lower watershed site (Mendon) the sources of discharge include reservoir releases and agricultural return flows, and the peaks in discharge, turbidity, and concentration are more attenuated than at Paradise. Additionally, concentration does not track discharge at Mendon as closely as it does at Paradise, indicating that sources of phosphorus at Mendon are not as closely related to discharge. The speciation of TP differs between the two sites. At Paradise, approximately 60 percent of the total phosphorus is in particulate form and 40 percent is dissolved whereas at Mendon, the ratio is reversed. Overall, there is more absolute variability in turbidity and resulting TP and TSS concentrations at Paradise than at Mendon, but there is greater short term variability at Mendon. The differing scales of Fig. 4.2 and 4.3 make it difficult to

compare short term variability between the two sites, but Fig. 4.4 shows turbidity at both sites for a three month period in late summer and early fall. For this time period, turbidity regularly fluctuates by 10-12 NTU within a day at Mendon while variations within a day at Paradise are on the order of 1.5-2 NTU.

The difference between the two water years (WY) examined in this paper should also be noted. WY 2006 was a relatively high flow year in the Little Bear due to a considerable snowpack and favorable conditions during runoff, while precipitation and discharge were both low in WY 2007.

4.3.2 Scenario generation and load estimation

The datasets of matched discharge and concentration at half hour intervals were decimated at varying frequencies to create subsets of paired discharge and concentration estimates from which annual loads were calculated. Equation 4.1 was used to calculate the load estimates for all of the subsets. This is a simple linear interpolation method, and is the most straightforward and accurate equation of the methods researched (de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Webb et al., 2000).

$$W = \sum_{i=0}^{n} Q_i C_i \, x \tag{4.1}$$

where W is the total annual load (kg), Q_i represents the incremental discharge (cms), C_i represents the incremental concentration of TP or TSS (mg/L), x is a factor to convert to kg per appropriate time period, and n is the total number of paired discharge and concentration estimates in one year (17520 for half hourly, 8760 for hourly, 365 for daily,

52 for weekly, and 12 for monthly). For all sampling frequencies, annual loads were calculated for WY 2006 and WY 2007. The subsets of data are described in Table 4.2.

4.3.2.1 High frequency

Using the complete sets of discharge and concentration data (half hourly), annual loads were calculated according to Equation 4.1. These values are the reference loads used for comparison with the other sampling frequencies. In order to test how much information was lost by sampling hourly instead of half hourly, a subset of discharge and concentration measured every hour was created, and annual loads were calculated.

4.3.2.2 Daily frequency

To represent sampling at a daily frequency, two types of subsets were generated. The first type was created by randomly selecting an instance of corresponding discharge and concentration within each day, resulting in 365 values per year. Equation 4.1 was then used to calculate annual loads. To achieve a distribution of load estimates using this method, random sampling and load calculation was conducted 10,000 times. The second type of daily subset was created to examine the effects of sampling time on load estimates. To simulate consistently sampling at the same hour of the day, corresponding discharge and concentration were selected for each hour of the day on every day of the year resulting in 24 subsets (one for each hour of the day) from which annual loads were calculated.

4.3.2.3 Weekly frequency

Two types of subsets were also generated to simulate weekly sampling. The first type was created by randomly selecting a single instance of corresponding discharge and concentration from within each week resulting in a decimated dataset with one discharge and concentration for each week (52 values for each year). This was conducted 10,000 times, and 10,000 annual loads were subsequently calculated. The second type of weekly sampling was designed to assess the impact of consistently sampling on a particular day of the week. Corresponding values of discharge and concentration were randomly selected from one day of the week for an entire year, resulting in 52 values of paired concentration and discharge (one for each week of the year) from which an annual load was calculated. In order to obtain a distribution of results using this method, random selection and load calculation was conducted 10,000 times. This procedure was repeated for each day of the week, resulting in a total of 70,000 annual load calculations.

4.3.2.4 Monthly frequency

Monthly sampling was simulated by randomly selecting a discharge and corresponding concentration within each calendar month resulting in 12 values for each year, from which annual loads were calculated. Ten thousand annual load calculations were realized.

4.4 Results

4.4.1 Frequency comparison

To illustrate the effects of sampling frequency, Fig. 4.5 shows series of TSS estimates at Paradise during spring runoff (February-May) of 2006. Conclusions can be extended to TP, both sites, and for longer time periods. The half hourly concentration time series is shown, along with subsets of the half hourly concentrations decimated at decreasing sampling frequency. The hourly series consists of concentrations on the hour, while the daily, weekly, and monthly series are randomly selected concentrations from the half hourly record. The hourly record shows little divergence from the half hourly dataset. The daily concentration record appears to capture the general trend of TSS concentration, but it fails to portray the fine resolution variability. The weekly and monthly series completely miss the peaks in concentration, which are the periods of greatest contribution to total annual load. On the other hand, under a monthly or weekly sampling routine, a sample could be collected during a peak in concentration leading to a significant overestimation of annual load.

Scholefield et al. (2005) recommend that the sampling frequency should match the scale of the processes involved. Kirchner et al. (2004) assert that the measurement frequency of chemical constituents should be often enough that no new information is gained by sampling more frequently. In this case, the half hourly concentrations do not reveal any pattern that is not observed in the hourly data, but the daily concentrations overlook behavior that is occurring within the day. For other watersheds or other constituents, making measurements more frequently than hourly or half hourly may be

necessary. For example, Tomlinson and De Carlo (2003) used in situ measures at five minute intervals to demonstrate the high variability in Hawaiian streams.

Table 4.3 and Fig. 4.6 and 4.7 summarize the results of load calculations for TP and TSS at Paradise and Mendon for WY 2006 and WY 2007. Fig. 4.6 and 4.7 include boxplots for all variables and years at Paradise and Mendon, respectively. The categories in the plots correspond to simulated sampling frequency including the reference loads (half hourly), hourly, randomized daily, randomized weekly, and randomized monthly. The boxes represent the first and third quartiles (25^{th} and 75^{th} percentiles) and the whiskers correspond to the lower and upper adjacent levels of the 10,000 realizations of annual load calculations. The medians of the 10,000 realizations of the randomized daily, weekly, and monthly subsets are also indicated. The percentage above the whisker represents the fraction of 10,000 realizations that fall above the upper adjacent level. There were no values below the lower adjacent levels. Table 4.3 summarizes the plots in Fig. 4.6 and 4.7 by reporting the bias, calculated with respect to the reference loads, of the lower and upper adjacent levels, the 1^{st} and 3^{rd} quartiles, and the median for each of the sampling frequencies.

At Paradise, for both variables and years, the median loads decrease as sampling frequency decreases, indicating that less frequent sampling typically omits periods of significant constituent loading and thus underestimates annual loads. This is consistent with the findings of Richards and Holloway (1987) and Phillips et al. (1999). In contrast, at Mendon, the median loads for all sampling frequencies are within 5 percent of the reference loads. Sampling frequency also affects the range of load estimates. For all

variables, sites, and years, the variability in the load increases as sampling frequency decreases because a single sample is assumed to be representative of a longer time period. The discharge and concentration at that point might not be characteristic of that time period (e.g., a sample collected during a rain on snowmelt event is assumed to represent an entire month), and the resulting annual load can be skewed.

Overall, hourly sampling frequency provides a very close approximation of the reference load at these sites, so little resolution is lost by decreasing sampling frequency to hourly. The departure of the load estimates from the reference load, as indicated by the bias calculations, varies between site and variable. In general, the loads at Mendon are closer to the reference loads than are those at Paradise. At a daily frequency for both variables and both years at Mendon, even the lower and upper adjacent levels are within 5 percent of the reference load. Additionally, at Mendon, the percentages of values falling above the upper adjacent level are all less than those observed at Paradise. At Paradise, TSS concentrations were more variable than TP as the medians for weekly and monthly sampling are all greater than 15 percent of the reference loads and the 1st and 3rd quartiles are not within 10 percent of the reference load for daily sampling. Richards and Holloway (1987) also found TSS to be more volatile than TP. No prominent difference between the two water years is observed apart from the differing scales as TP and TSS transport was greater in 2006 than in 2007.

4.4.2 Probability of achieving the reference loads

Although 10,000 load estimates were generated by randomly subsampling at daily, weekly, and monthly time scales, in reality, only one annual load estimate could be made using real sampling data, regardless of its frequency. Using the 10,000 load estimates for daily, weekly, and monthly frequencies, we examined the likelihood of a single load estimate falling within certain thresholds of the reference load. In other words, we asked how probable it is that we will be close to the true loading if we sample at the given frequency. Thresholds of 5 percent and 50 percent were selected to represent being very close to the reference load and being "within the ballpark" of the reference load, respectively. A few studies used 20 percent as an acceptable error from the reference load (Richards and Holloway, 1987; Coynel et al., 2004), but we think that more accurate loads are achievable. The results (reported in Table 4.4) further reveal differences between the two sampling sites. At Mendon, the probability of being within 5 percent of the reference load is 1.0 for sampling at a daily frequency, 0.50-0.75 for a weekly frequency, and 0.20-0.31 for monthly sampling. At Paradise, on the other hand, daily sampling only has a probability of 0.19-0.46 of achieving a load estimate within 5 percent of the reference load. At Mendon, it is very probable (0.98-1.0) that loads will be within 50 percent of the reference load, regardless of sampling frequency. In contrast, with monthly sampling at Paradise, the probability of being within 50 percent of the reference load is only 0.52-0.89.

4.4.3 Daily by hour loads

The variability in loads calculated by simulating consistently sampling at the same time each day is shown in Fig. 4.8. Though the trends are distinct for each site, they are similar across variables and years. At Paradise, loads calculated from concentrations and discharges at the end of the day (hours 16-24) are higher than those calculated for hours earlier in the day, although the increase is less dramatic for both TP and TSS in WY 2007. At the most extreme, the loads vary by 50 percent from sampling at one hour as opposed to another hour of the day. At Mendon, the highest loads are in the early hours of the morning (hours 2-6), but overall, there is less variability throughout the day than at Paradise.

We believe that the differences in loads throughout the day are due to diurnal fluctuations in turbidity (and resulting TP and TSS), as shown in Fig. 4.4. Limited grab sampling at Paradise reveals a broad range of TP and TSS values within a single day. During the height of spring runoff, 24-hour sampling was conducted returning TP concentrations ranging from 0.066 and 0.954 mg/L and TSS ranging from 108 to 2450 mg/L.

The site specific hydrologic conditions are probable causes of the differing patterns between the two sites. For example, the timing of the response to events such as snowmelt or storms varies between the upper and lower watersheds. The timing of reservoir releases may also affect the timing of loads at Mendon. Additional factors that could cause varying behavior within a day include changes in water temperature, evapotranspiration, and the timing of agricultural withdrawals. Scholefied et al. (2005)

suggest that diurnal fluctuations in phosphorus concentrations may be a result of enrichment or depletion by instream biological processes or physical processes that are a response to temperature. The authors point out that if the diurnal variations are a result of physical processes, the intensity of variation will decrease in a downstream direction, which is the case with Paradise and Mendon. Jordan et al. (2007) attribute diurnal phosphorus fluctuations to rural point sources upstream of the sampling site, which may also explain some of the diurnal variability at both locations.

4.4.4 Weekly by day loads

Fig. 4.9 is a collection of boxplots representing annual loads that simulate sampling once a week on the same day each week but randomizing the time of day of sampling. The trends vary between sites as well as between water years. At Paradise, in WY 2006, both TP and TSS loads calculated from sampling on Tuesdays and Wednesdays were consistently higher and exhibited greater variability than other days of the week. In contrast, loads for WY 2007 at Paradise were more consistent and exhibited less variability for all days of the week. At Mendon, there is no obvious pattern in loads or variability based on day of the week, water year, or variable. Although no trend is observed, the ranges of loads are still notably different between different days of the week. These results indicate the day of the week that sampling is conducted impacts the load estimate, but we have little rationale for the observed patterns. One possibility is that the differences between days of the week are related to the days on which diversions are opened or closed.

4.4.5 Site comparison

These results demonstrate differences between Paradise, the upper watershed site, and Mendon, the lower watershed site. As mentioned, at Mendon, the river generally has higher discharge, is lower gradient, and has more interaction with groundwater. Paradise, on the other hand, has lower baseflow and higher peaks in discharge due to the higher gradient of the river as well as the surrounding land. In general, the results show that high frequency sampling is essential for load calculation at Paradise, but that, depending on the level of acceptable error, less frequent sampling can be conducted at Mendon. This is consistent with the findings of a number of studies comparing sampling frequency on different rivers. As smaller rivers are more responsive to precipitation and snowmelt while the responses of larger rivers are more attenuated with a slower rise in discharge and higher baseflow levels (Richards and Holloway, 1987), there is a greater decrease in precision of load estimate with reduced sampling frequency for smaller rivers (Richards and Holloway, 1987; de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Phillips et al., 1999; Coynel et al., 2004). Also, rivers with high baseflow in permeable lowlands have less variable TP and TSS, so a lower sampling frequency is acceptable compared to rivers with low baseflow that transport more TP and TSS in high discharge events and require more frequent sampling (Ferguson, 1987; Johnes, 2007). In addition to a greater bias in load estimates at Paradise than Mendon, there was a greater degree of underestimation of loads at Paradise than at Mendon. This is consistent with studies that found that smaller rivers tend to underestimate loads more so than larger rivers (Kronvang and Bruhn, 1996; Phillips et al., 1999). Although Paradise and Mendon are

located on the same river, the behavior of the river changes dramatically between the two sites and Paradise can be seen as a small river site while the attributes of a larger river could be ascribed to the Little Bear River at Mendon.

4.5 Conclusions

This paper used high frequency records of TSS and TP concentrations, which were estimated using surrogate relationships with turbidity, along with matched series of discharge to calculate reference loads for two sites on the Little Bear River. Two water years (2006 and 2007) of data were used. WY 2006 had high discharge (and high constituent transport) relative to WY 2007, which was a low discharge year.

In order to simulate decreasing sampling frequencies, the continuous records were decimated at hourly, daily, weekly, and monthly intervals, and annual loads were calculated from the resulting subsets. For the daily, weekly, and monthly frequency subsets, a single value of discharge and a corresponding concentration were selected at random from within the sampling period and subsequently used to calculate annual loads. At each sampling frequency for each variable and each water year at each site, 10,000 realizations of annual load were generated for the purpose of examining the potential variability in annual load estimates.

The hourly loads were a close approximation of the reference loads across sites, variables, and years. At the upper site, sampling with decreased frequency resulted in median annual loads that were increasingly less than the reference loads as important periods in TP and TSS transport were overlooked. For both sites, decreasing sampling frequency increased the variability in the load calculations as there is a high probability

of using a set of discharge and concentration records that are not representative of the entire time period. The distribution of annual loads varied between sites as well as variables. There was a much greater variation in loads at Paradise than at Mendon, which may be attributed to the different hydrologic characteristics of the two sites as well as differences in TP and TSS sources and behavior.

In addition to annual loads determined by randomly sampling at daily, weekly and monthly frequencies, loads were also calculated to simulate sampling at the same time every day and the same day every week. Overall, the results show that the time of day and the day of week that sampling is conducted have a substantial impact on annual load calculations, although in this case, the level of impact varies between site and year.

We conclude that periodic grab sampling, even at a daily frequency, is not a suitable substitute for loads calculated from continuously estimated concentrations. Although loads calculated from subsampling at a daily frequency may have low variability from the reference loads, depending on the site, the hour of the day on which sampling is collected has a significant impact on load estimates. Additionally, daily sampling for extended time periods is cost prohibitive and logistically difficult. Hourly or half hourly measurements capture the fluctuations that occur in concentration and discharge at a finer scale than daily data can achieve. Furthermore, the loads calculated from weekly and monthly subsampling do not adequately approximate the reference loads, and caution should be taken in calculating loads from data at this sampling frequency. However, the degree of variability depends on the site and the variable.

Using high frequency surrogates to calculate constituent loads overcomes many of the inadequacies of loads estimated from periodic grab sampling as it provides increased resolution and accuracy while remaining logistically and economically feasible. High frequency, in situ monitoring with surrogate relationships for concentration should be considered as a representative and economically feasible alternative to periodic grab sampling for load calculations. High frequency measurements will provide water quality monitoring programs, regulatory agencies, and environmental observatories with an improved view of constituent behavior.

Table 4.1
Relationships used to derive continuous concentration time series

Site	Constituent	Equation
Paradise	Total Phosphorus	$TP = 0.209 + 0.000798 * Turb + 0.0386 * Z$
	Total Suspended Solids	$TSS = 3.58 + 1.31 * Turb$
Mendon	Total Phosphorus	$TP = -0.0341 + 0.0053 * Turb + 0.0949 * Z - 0.00404 * Turb * Z + 0.0832 * Y - 0.00871 * Y * Turb$
	Total Suspended Solids	$TSS = 0.341 + 1.41 * Turb$

Variable	Description
TP	Total Phosphorus, mg/L
TSS	Total Suspended Solids, mg/L
Turb	Turbidity, NTU
Z	categorical variable for spring runoff ($Z = 1$) versus baseflow ($Z = 0$)
Y	categorical variable for *Turb* <10 NTU ($Y = 1$) versus *Turb* >10 NTU ($Y = 0$)

Table 4.2
Summary of decimated datasets

Subset	Frequency	Realizations
Complete	Half Hourly	1
Hourly	Hourly	1
Daily by Hour	Daily	24*
Randomized Daily	Daily	10,000
Weekly by Day	Weekly	70,000**
Randomized Weekly	Weekly	10,000
Randomized Monthly	Monthly	10,000

*one realization was generated for each hour of the day
**10,000 realizations were generated for each day of the week

Table 4.3
Biases (percentages) of the ranges of load estimates based on reference loads. Biases within 5 percent of the reference load are highlighted

Site	Variable	Year	Frequency	Lower Adjacent	1st Quartile	Median	3rd Quartile	Upper Adjacent
Paradise	TP	2006	Hourly			0.003		
			Daily	-24	-8.2	-1.2	6.7	29
			Weekly	-43	-24	-11	10	61
			Monthly	-59	-37	-25	-1.8	50
	TP	2007	Hourly			-0.01		
			Daily	-16	-6.1	-1.2	4.6	21
			Weekly	-27	-15	-8.6	6.3	38
			Monthly	-59	-26	-13	5.1	50
	TSS	2006	Hourly			-0.008		
			Daily	-40	-14	-1.6	12	51
			Weekly	-69	-41	-19	14	97
			Monthly	-81	-60	-46	-13	57
	TSS	2007	Hourly			-0.04		
			Daily	-40	-15	-2.7	12	53
			Weekly	-57	-38	-22	13	90
			Monthly	-76	-55	-40	-6.0	65
Mendon	TP	2006	Hourly			-0.004		
			Daily	-3.7	-0.88	-0.17	0.64	2.9
			Weekly	-14	-3.5	-0.43	2.7	12
			Monthly	-27	-10	-3.1	5.0	27
	TP	2007	Hourly			0.04		
			Daily	-4.6	-1.1	-0.18	0.75	3.5
			Weekly	-14	-4.1	-0.77	2.8	13
			Monthly	-38	-11	-1.4	9.99	40
	TSS	2006	Hourly			0.01		
			Daily	-4.6	-1.2	-0.22	0.75	3.6
			Weekly	-15	-3.9	-0.23	3.3	14
			Monthly	-36	-12	-3.0	8.26	39
	TSS	2007	Hourly			0.04		
			Daily	-5.63	-1.7	-0.43	0.98	5.0
			Weekly	-24	-5.3	-0.60	4.69	20
			Monthly	-49	-14	-1.8	12	47

Table 4.4
Probabilities of falling within a certain threshold of the reference load. Probabilities are determined by using the 10,000 realizations of random load calculations

			Probability of being within 5% of the reference load			Probability of being within 50% of the reference load		
Site	Variable	Year	Sampling Frequency					
			Daily	Weekly	Monthly	Daily	Weekly	Monthly
Paradise	TP	2006	0.33	0.12	0.056	1.0	0.90	0.85
	TP	2007	0.46	0.15	0.14	1.0	0.95	0.89
	TSS	2006	0.19	0.071	0.022	0.99	0.72	0.43
	TSS	2007	0.19	0.057	0.044	0.99	0.82	0.52
Mendon	TP	2006	1.0	0.72	0.31	1.0	1.0	0.99
	TP	2007	1.0	0.67	0.25	1.0	1.0	0.99
	TSS	2006	1.0	0.66	0.23	1.0	1.0	0.99
	TSS	2007	1.0	0.50	0.20	1.0	1.0	0.98

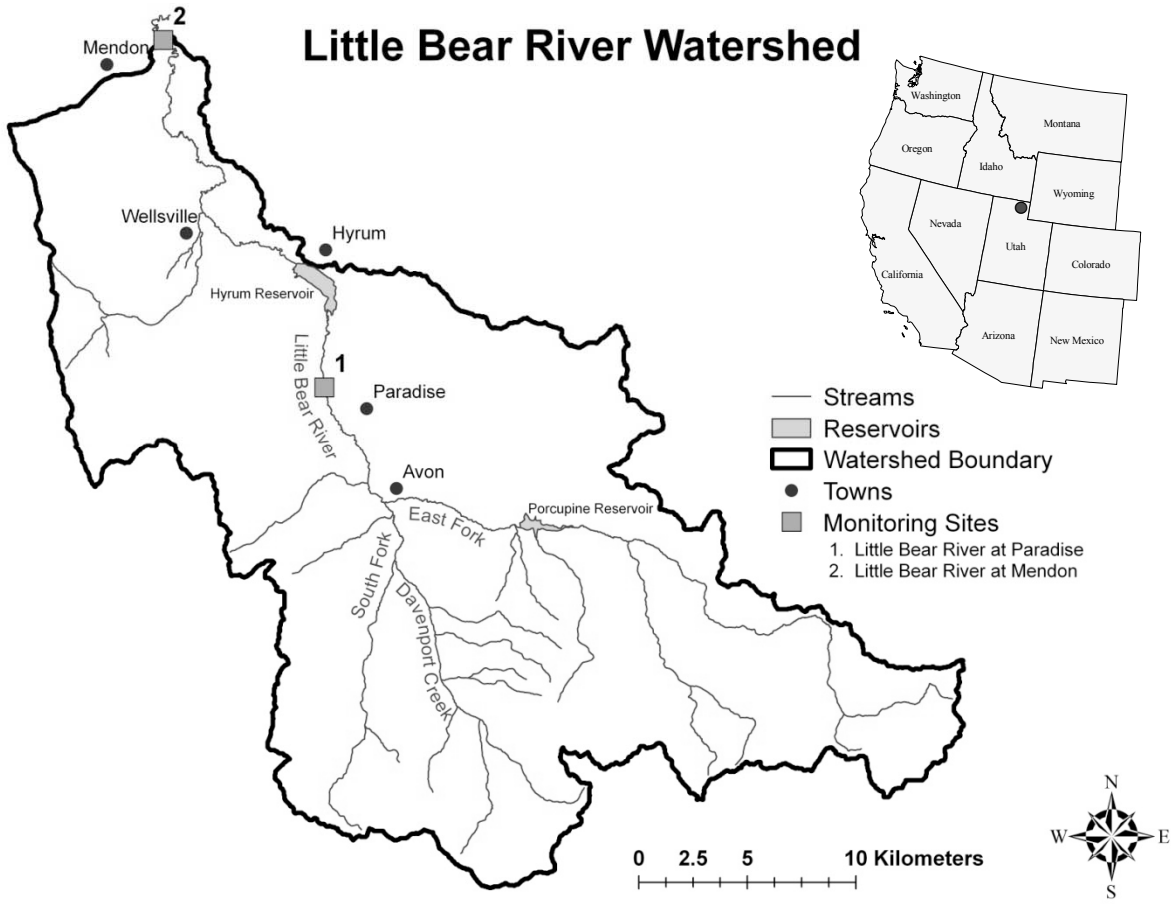

Fig. 4.1. Little Bear River watershed.

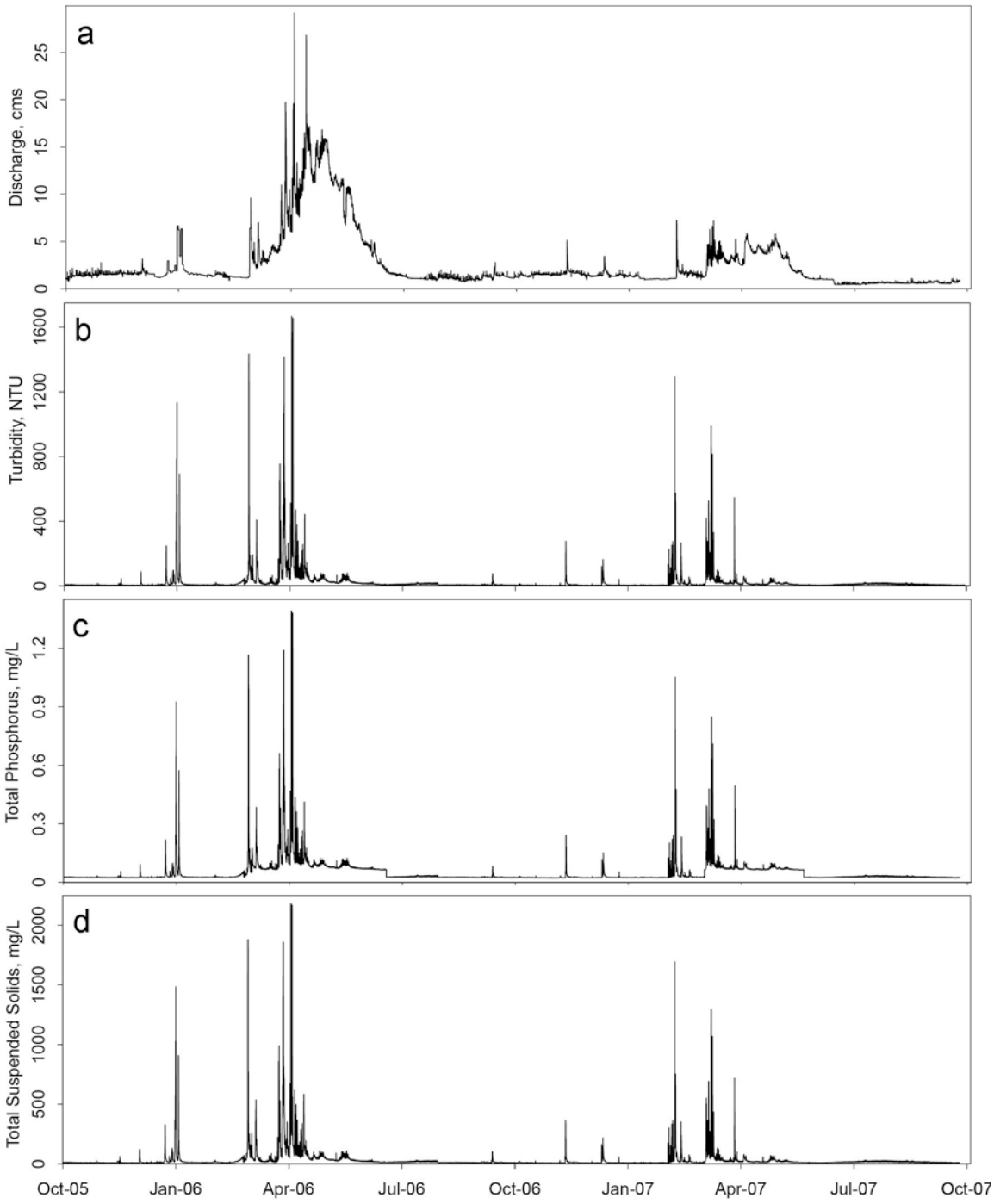

Fig. 4.2. Half hourly discharge (a), turbidity (b), TP concentration (c), and TSS concentration (d) at Paradise.

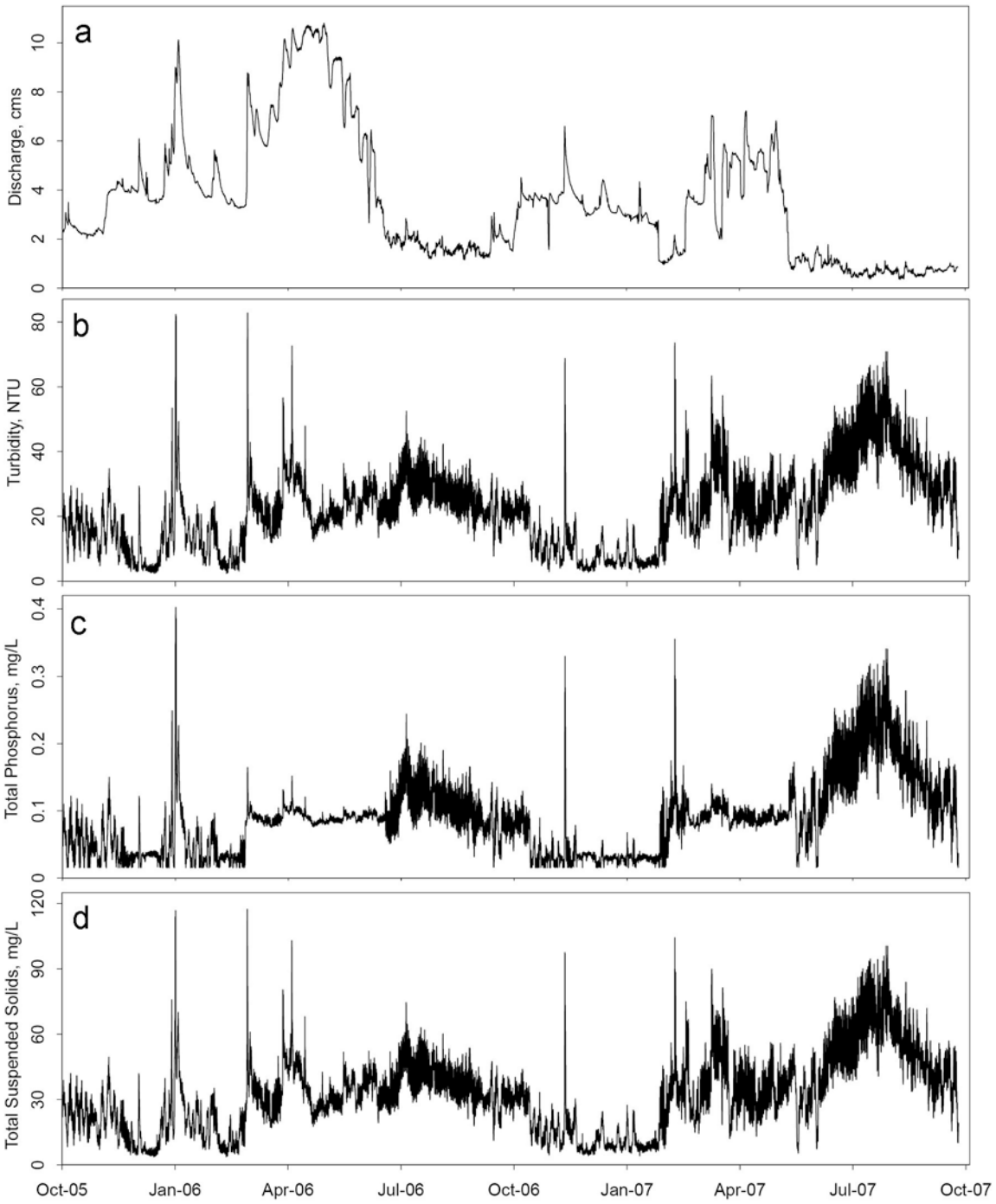

Fig. 4.3. Half hourly discharge (a), turbidity (b), TP concentration (c), and TSS concentration (d) at Mendon.

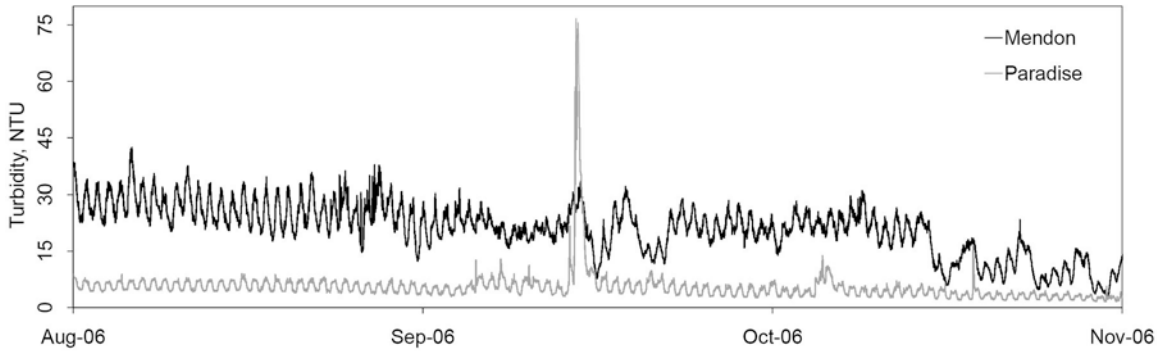

Fig. 4.4. Turbidity at Paradise and Mendon.

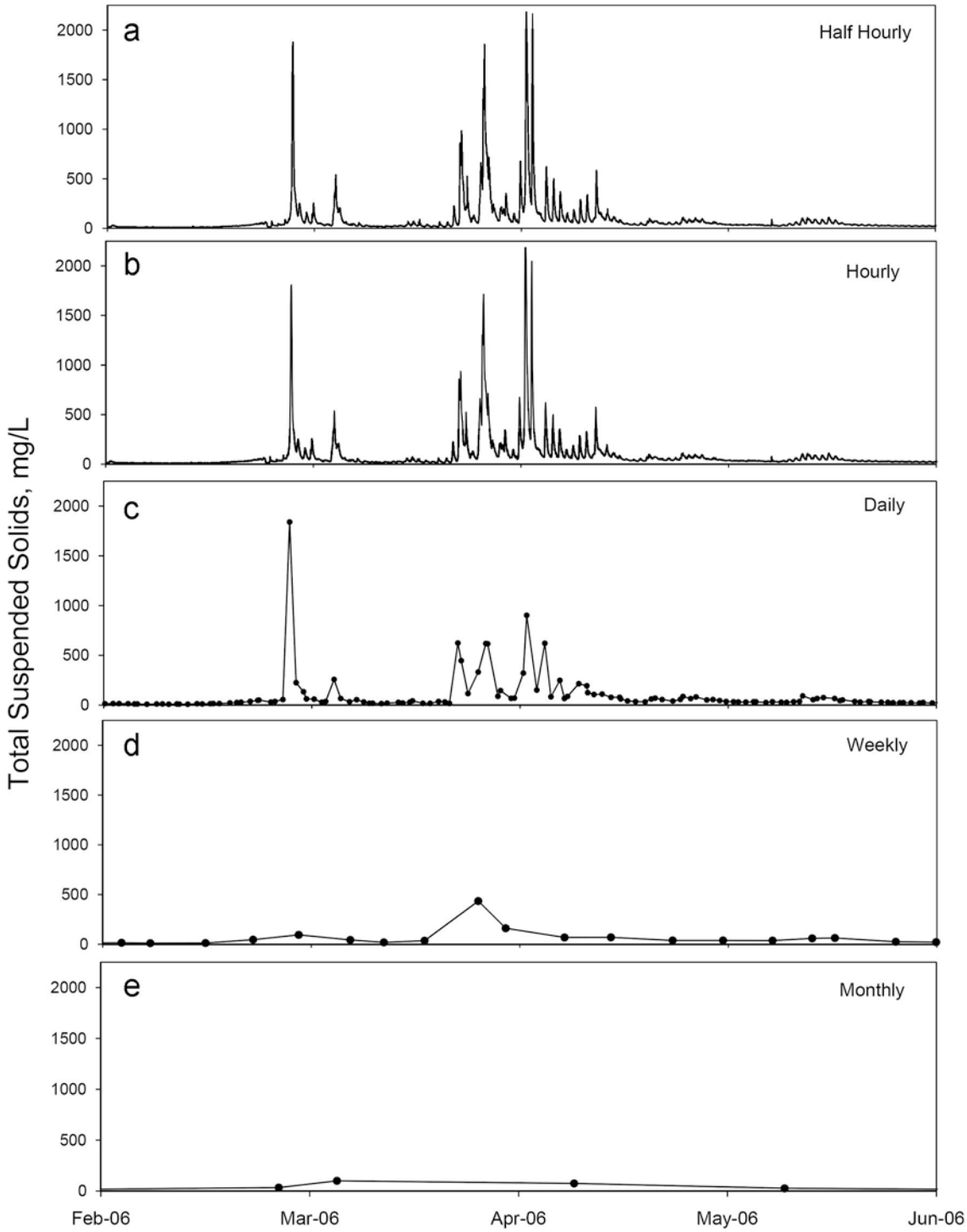

Fig. 4.5. TSS concentrations at varying sampling frequencies as subsampled from the half hourly concentration estimates (a). The hourly time series (b) consists of estimates made on the hour while the daily (c), weekly (d), and monthly (e) are randomly selected points.

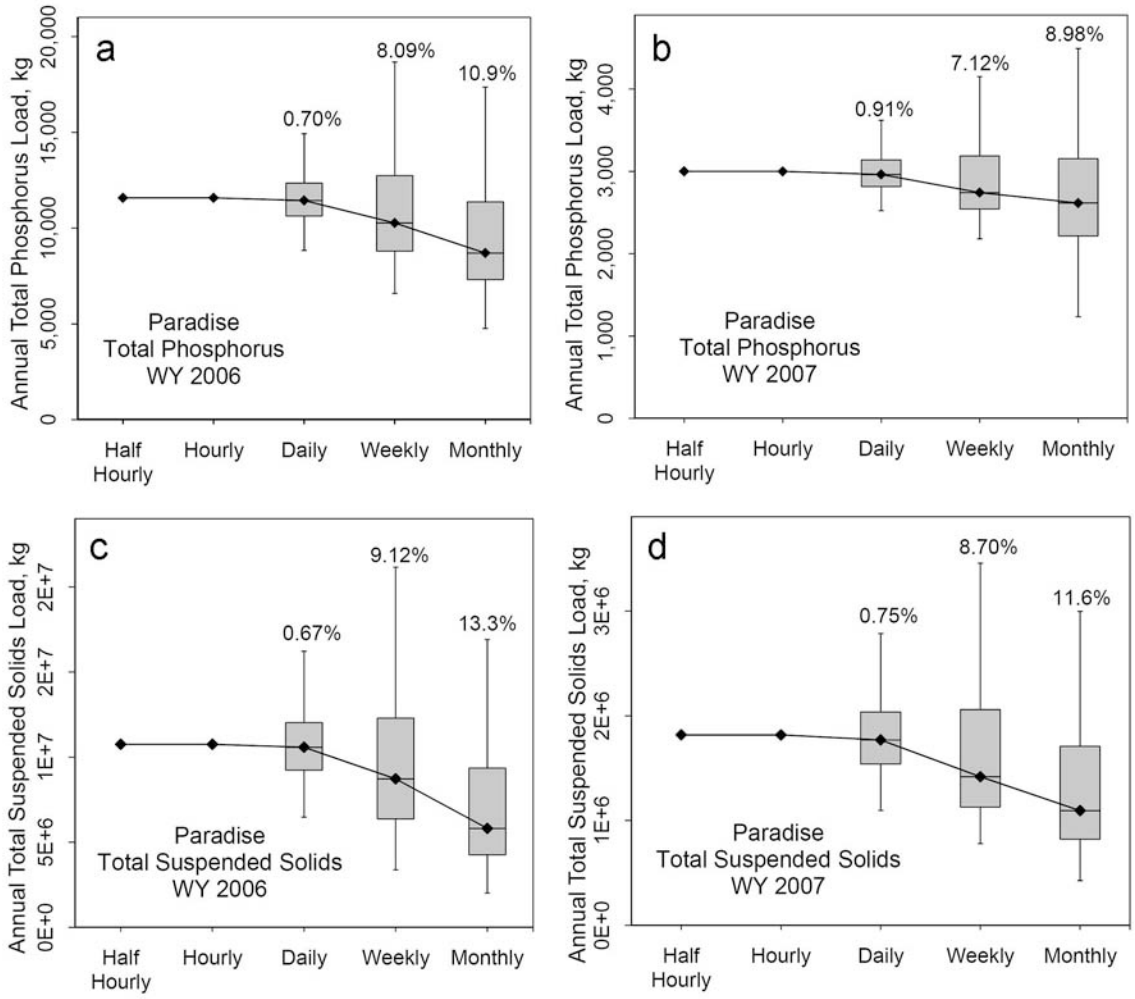

Fig. 4.6. Box and whisker plots of the results of varying sampling frequencies at Paradise. Half hourly represents the reference load, hourly represents the value from subsampling on the hour, and daily, hourly, and monthly represent 10,000 realizations of randomly selected points within each day, week or month. The boxes represent the first and third quartiles and the whiskers represent the lower and upper adjacent levels. The medians of each of the sets of realizations are also indicated. The percentages above the upper whisker represent the portion of calculated loads that fell above the upper adjacent level. There were no values below the lower adjacent levels.

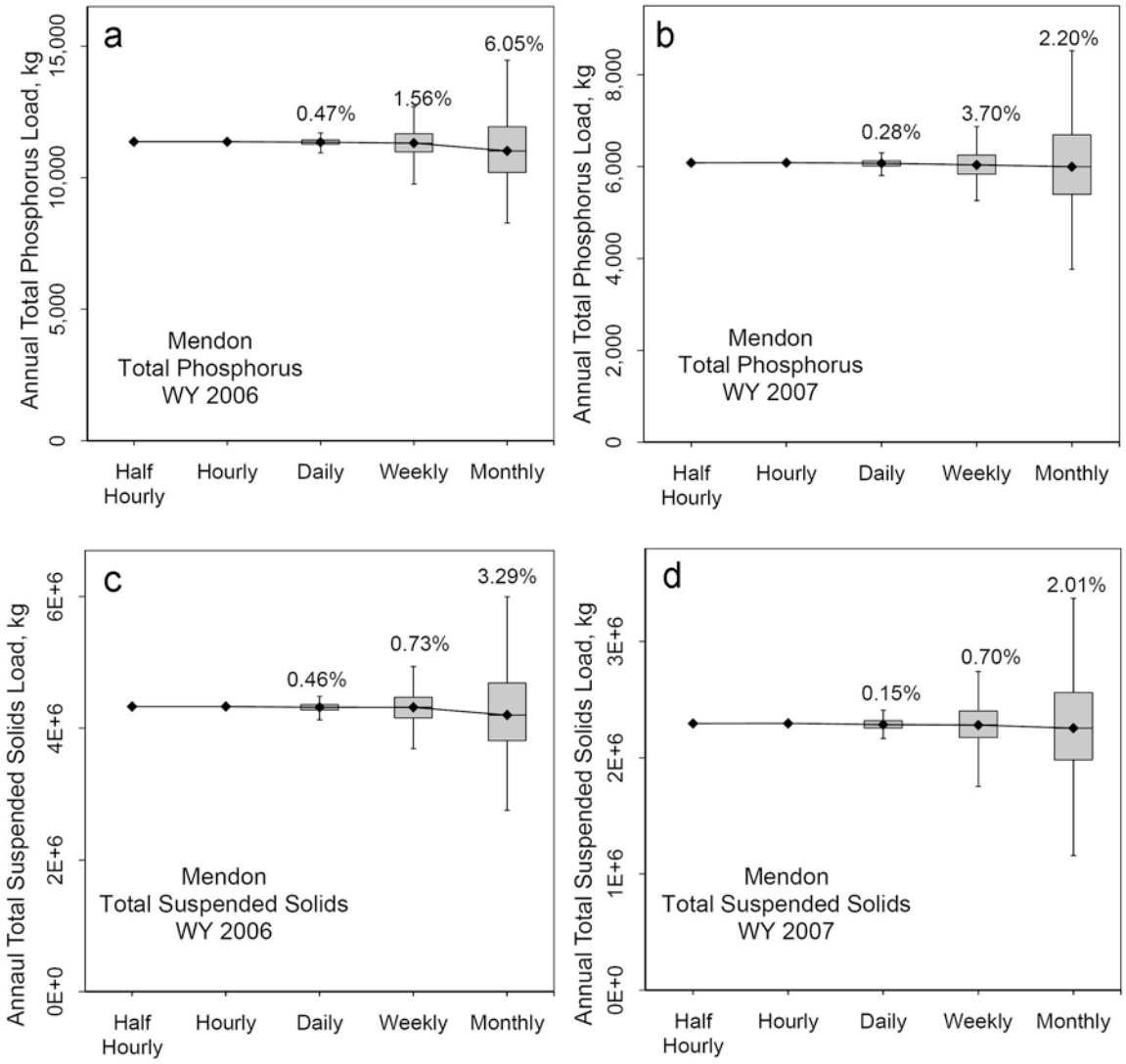

Fig. 4.7. Box and whisker plots of the results of varying sampling frequencies at Mendon. See Fig. 4.6 for a description of the boxes and whiskers

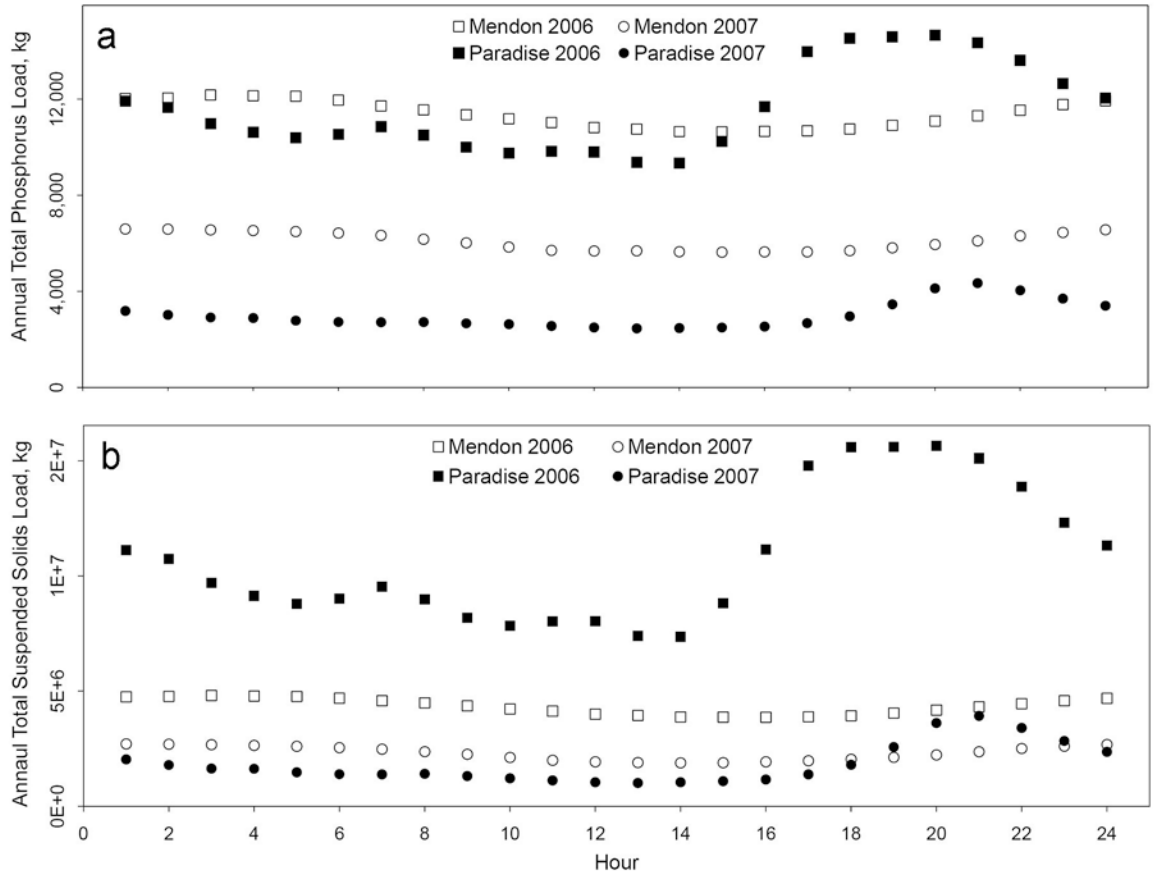

Fig. 4.8. Annual loads calculated by subsampling daily at the same hour each day.

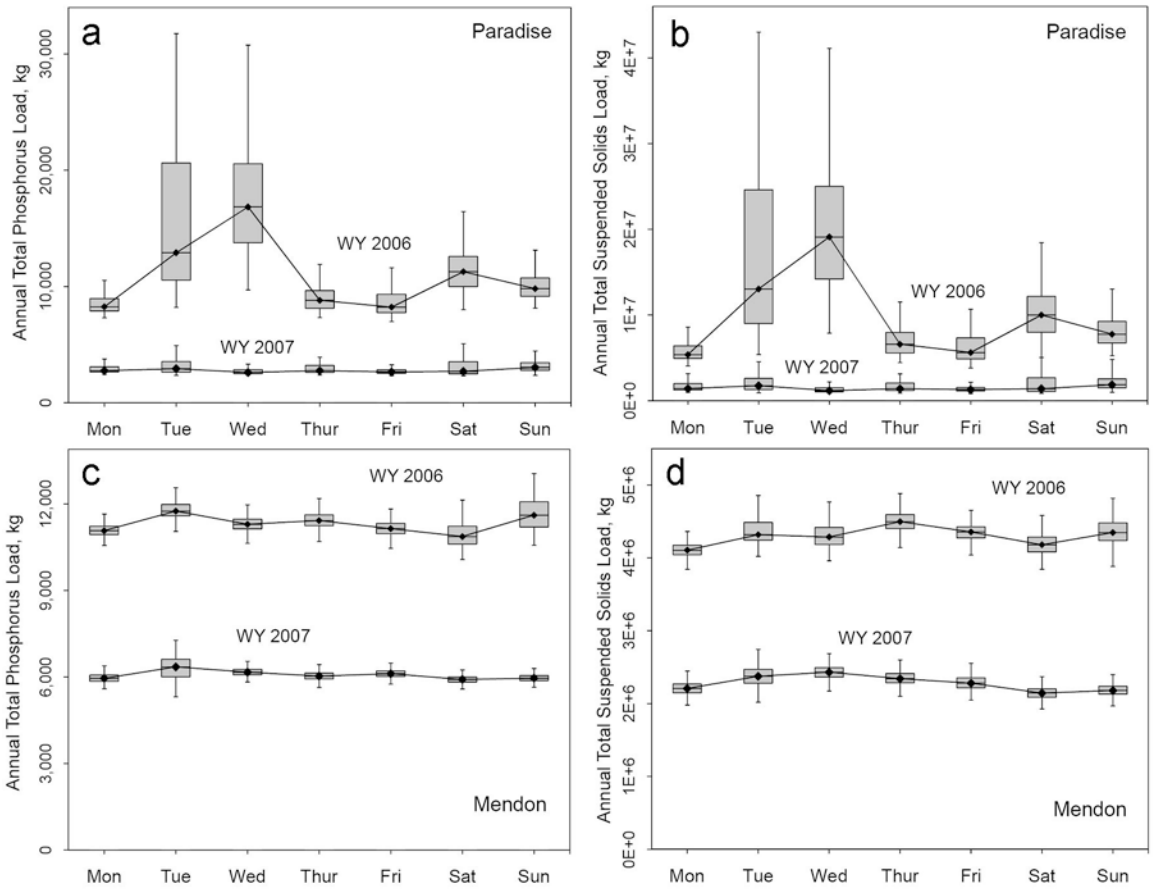

Fig. 4.9. Box and whisker plots of annual loads calculated by consistently subsampling on the same day of the week. Sampling frequency is weekly using a randomly selected point within a single day. Statistics are based on 10,000 realizations of annual load for each day, each site, each variable, and each water year. See Fig. 4.6 for a description of the boxes and whiskers.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The uncertainty in concentration trends and associated load calculations resulting from low sampling frequency is a pressing challenge for water quality monitoring programs. Various researchers have investigated techniques to address deficiencies presented by the sparse datasets generated by traditional water quality monitoring. One of these approaches involves complex equations for load calculation to account for sporadic sample collection, but no single estimation method has been deemed appropriate for all watersheds, constituents, hydrologic patterns, and sampling frequencies. Another approach to overcoming the limitations of sampling frequency is using in situ sensors to continuously measure water quality constituents. This research presented in situ turbidity as a surrogate measure for total phosphorus (TP) and total suspended solids (TSS) as an alternative to intermittent grab sampling. To examine the effects of sampling frequency on load calculations, high frequency concentration estimates were generated from the surrogate relationships and subsampled to simulate decreasing sampling frequency.

The surrogate relationships were generated at two locations on the Little Bear River using two years of high frequency turbidity data and intermittently sampled TP and TSS. In order to account for the number of censored data points within the TP datasets, maximum likelihood regression within the statistical program R was used to generate the parameters for regression equations for TP. The basic linear regression function in R was used to generate the parameters for TSS. Additional explanatory variables examined were discharge, water temperature, day of year, hour of day, and categorical variables

representing spring snowmelt runoff versus baseflow conditions and the occurrence of storm events. At both sites, turbidity and the categorical variable representing runoff/baseflow were the only significant explanatory variables for TP, indicating that the relationships between turbidity and TP are consistent throughout storm events. However, the regression at the lower site was greatly improved by the inclusion of the interaction between turbidity and the runoff/baseflow categorical variable, and an additional categorical variable was necessary for low turbidity conditions. For TSS, turbidity was the only significant explanatory variable at both sites, indicating that the relationship between turbidity and TSS is consistent across hydrological conditions. Logarithmic transformations of the datasets did not provide any improvement in the models.

Using the root mean square error as an estimation of overall error in the regression equations, all of the relationships had error values of one-fourth to one-half of the mean of the observed data. Visual examinations of the observed and estimated concentrations indicate that the equations adequately track observed trends.

The surrogate relationships were used with the continuously collected turbidity data to generate high frequency estimates of TP and TSS concentration. Along with high frequency estimates of discharge, the concentration data were used to calculate annual loads of TP and TSS for two water years, creating reference loads. In order to examine the effect of sampling frequency on load estimation, the concentration and discharge records were decimated at hourly, daily, weekly, and monthly intervals to represent grab sampling at those frequencies. Annual loads were calculated from the decimated datasets and compared to the reference loads. For the daily, weekly, and monthly datasets,

concentration and corresponding discharge were randomly selected 10,000 times to generate a distribution of annual load estimates.

Loads calculated from the hourly concentration and discharge closely approximated the reference loads. For both TP and TSS, at both sites, and for both water years, the variability in annual load estimates increased as sampling frequency decreased because a single point of concentration and discharge was assumed to represent an extended time period. At the upper watershed site, however, the variability was greater. Also, at the upper site, the median loads consistently decreased as sampling frequency decreased, verifying that intermittent sampling omits important periods of constituent transport and generally underestimates annual loads at this location. The levels of bias from the reference load differed between sites and variables, but were fairly consistent for the two water years examined. There was more bias and variability in loads estimated at the upper watershed site than the lower watershed site, and at the upper site, TSS loads were more biased than TP loads. The probability of calculating loads within certain thresholds of the reference loads was also examined. The results show a greater probability of approaching the reference load at higher sampling frequencies. Furthermore, the probability of approximating the reference loads was greater at the lower watershed site. The differences in hydrologic response as well as TSS and TP behavior at the two sites are thought to explain the differing results.

The timing of sample collection was also examined. Annual loads were calculated by subsampling at the same time each day as well as subsampling on the same day of the week. Results indicate the time of day of sample collection has an impact on

resulting loads. The pattern differed from one site to another, reflecting diurnal fluctuation in turbidity and TP and TSS concentrations likely due to the timing of hydrologic response. The degree of variability in loads calculated at different times of the day was different between two water years. Consistently sampling on the same day of the week also affects load estimates, depending on the site, the variable, and the water year.

This research has demonstrated the powerful potential of surrogate measures for generating high frequency concentration estimates from which loads can be calculated. The datasets generated by surrogate relationships provide information showing the high resolution dynamics of constituents that could not be attained using monthly, weekly, or even daily grab sampled concentration. Conventional grab sampling is also insufficient for load calculations as it can severely under or over estimate annual loads. Surrogate measures can provide high frequency estimates of concentration over extended periods of time and at multiple locations, allowing for better understanding of constituent fluxes throughout the watershed and throughout hydrological conditions. For some sites and some variables, daily sampling may provide a reasonable estimate of annual load, but daily grab sampling for extended periods of time and at many sites is generally impractical. Until in situ technology is developed to viably measure important constituents such as TP and TSS, surrogate measures provide an economically and logistically feasible method for quantifying constituent flux at a high frequency over large temporal and spatial scales.

CHAPTER 6

ENGINEERING SIGNIFICANCE

In the fields of environmental engineering and water resources, there is a need for improved understanding and prediction of short and long term behavior of instream processes. For many constituents at many locations, instream variability occurs on a time scale of minutes or hours, not weeks or months, the frequency at which traditional water quality programs have conducted monitoring. This research demonstrates the value of surrogate measures to estimate water quality constituents. Surrogate measures can significantly increase the resolution of available concentration data over multiple years and at multiple locations throughout a study area.

Surrogate measures have implications for water quality monitoring and compliance, watershed studies, water quality modeling, and environmental observatory design. Additional benefits to in situ sensors that monitor continuously include automated data collection, the ability to connect data to a water quality model or to the Internet, the minimization of human errors and time delays, and an overall reduction in the cost of monitoring.

The widespread incorporation of surrogate measures into water quality monitoring programs will allow for the characterization of fluxes from one site to another along a river or between tributaries to a common lake or reservoir or from one type of terrain to another. Additionally, comparisons can be made between varying time scales such as the response to two different storm events, different behavior during spring

snowmelt in a high discharge year opposed to a drought year, or the daily or annual effects of reservoir releases.

Compliance with water quality standards is often based on a concentration or a load threshold. Loads determined by high frequency concentrations calculated from surrogate measures will allow for the determination of compliance with increased certainty. Additionally, the high resolution of concentrations estimated by surrogate measures will assist in the determination of compliance based on peaks and duration of concentration. Sensors for making surrogate measures can be installed at locations other than rivers and streams where water quality is a concern such as beaches, lakes, and wastewater treatment plants.

Many hydrologic and water quality models require extensive parameterization in order to predict water quality given changes in land use, management practices, or hydrological conditions. These parameters are calibrated for streams and watersheds using water quality observations. Concentrations from surrogate measures will provide an increased number of observations so that model parameters can be determined with more certainty.

Environmental observatories have received attention as settings where improved understanding of hydrologic and water quality processes can occur as they generate data at high temporal frequencies and high spatial densities. Surrogate measures are necessary to the design of environmental observatories because they provide a relatively inexpensive and logistically viable method for determining concentrations of constituents that cannot be measured in situ.

CHAPTER 7

RECOMMENDATIONS FOR FUTURE RESEARCH

A number of ideas that were identified as additional topics of research stemming from the generation of surrogate relationships and examination of sampling frequency of total phosphorus (TP) and total suspended solids (TSS) on the Little Bear River.

1. Since the beginning of this study, five sites in addition to Paradise and Mendon have been instrumented with continuous monitoring equipment. When a sufficient number of concentration measurements have been made at these sites (the current number of observations is on the order of 50 at each site over a period of 6-8 months), techniques similar to those described in Chapter 3 should be used to generate surrogate relationships for TP and TSS for each monitoring site. Relationships at additional sites may provide increased understanding of the behavior of the constituents throughout the watershed.
2. In addition to turbidity, water level, and water temperature, all of the sites have been instrumented with sensors to monitor pH, specific conductance, and dissolved oxygen. These variables should be investigated as potential explanatory variables for TP and TSS. Because specific conductance and pH are related to dissolved species, they may be especially valuable at Mendon where the majority of TP is dissolved. These variables could also be explored as potential explanatory variables for dissolved total phosphorus at all sites since dissolved phosphorus is biologically important.

3. Since the development of the surrogate relationships described in Chapter 3, additional TP and TSS samples have been collected and analyzed. These data should be used to corroborate the current models and to further refine the equations.
4. To this point, phosphorus has been considered the limiting nutrient in Cutler Reservoir and the Little Bear River. If nitrogen is determined to be a limiting nutrient, then sample analysis should include species of nitrogen, and surrogate relationships should be developed for nitrogen species.
5. The distinction between spring runoff and baseflow was important for the TP surrogate relationships, however, in the resulting concentration estimates, there are distinct steps when the transitions between baseflow and spring runoff occur. Instead of a categorical variable with a value of 0 or 1, a continuous variable could be incorporated that represents the percent of discharge that is runoff. The values of the variable could be determined using baseflow separation techniques.
6. In addition to in situ measures and variables representing hydrological conditions, variables corresponding to land use and watershed attributes could be more directly incorporated into the surrogate relationships. For example, if one of the variables was percent agricultural land above the monitoring site, this variable could be adjusted to simulate a management practice. Other variables that could be examined include average slope, contributing area, soil moisture capacity, and percent area of various land uses.

7. This research presented surrogate measures as an economical alternative to conventional grab sampling. It would be valuable to do a complete cost-benefit analysis comparing continuous monitoring to periodic grab sampling at multiple sampling frequencies. The expenditures of continuous monitoring include sensors, telemetry equipment, materials needed for installation, supplies for sample collection, sample analysis, and the cost of personnel for site maintenance and sample collection. The expenses of periodic grab sampling include supplies for sample collection, sample analysis, and the cost of personnel for sample collection.
8. Chapter 4 compared continuous monitoring to less frequent grab sampling, but there is little guidance regarding the amount, timing, and frequency of samples that should be collected in order to develop and maintain surrogate relationships. Investigation into this question would assist in environmental observatory planning.

REFERENCES

Berthouex PM, Brown LC. Statistics for environmental engineers. New York: Lewis Publishers, 2002.

Brasington J, Richards K. Turbidity and suspended sediment dynamics in small catchments in the Nepal Middle Hills. Hydrol. Processes 2000; 14: 2559-2574.

Burkart MR, Simpkins WW, Morrow AJ, Gannon JM. Occurrence of total dissolved phosphorus in unconsolidated aquifers and aquitards in Iowa. J. Am. Water Resour. Assoc. 2004; 40: 827-834.

Chapman T. A comparison of algorithms for stream flow recession and baseflow separation. Hydrol. Processes 1999; 13: 701-714.

Christensen VG. Characterization of surface-water quality based on real-time monitoring and regression analysis, Quivira National Wildlife Refuge, south-central Kansas, December 1998 through June 2001. Water Resources Investigations Report. USGS, 2001, pp. 28.

Christensen VG, Jian X, Ziegler AC. Regression analysis and real-time water-quality monitoring to estimate constituent concentrations, loads, and yields in the Little Arkansas River, south-central Kansas, 1995-99. Water Resources Investigations Report. USGS, 2000, pp. 36.

Christensen VG, Rasmussen PP, Ziegler AC. Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams. Water Sci. Technol. 2002; 45: 205-211.

Correll DL, Jordan TE, Weller DE. Transport of nitrogen and phosphorus from Rhode River watersheds during storm events. Water Resour. Res. 1999; 35: 2513-2521.

Coynel A, Schafer J, Hurtrez JE, Dumas J, Etcheber H, Blanc G. Sampling frequency and accuracy of SPM flux estimates in two contrasted drainage basins. Sci. Total Environment 2004; 330: 233-247.

Croke BFW, Jakeman AJ. Predictions in catchment hydrology: an Australian perspective. Marine and Freshwater Res. 2001; 52: 65-79.

de Vries A, Klavers HC. Riverine fluxes of pollutants: monitoring strategy first, calculation methods second. European Water Poll. Control 1994; 4: 12-17.

Etchells T, Tan KS, Fox D. Quantifying the uncertainty of nutrient load estimates in the Shepparton Irrigation Region. Australian Centre for Environmetrics, University of Melbourne, http://www.mssanz.org.au/modsim05/papers/etchells_2.pdf, 2005.

Ferguson RI. Accuracy and precision of methods for estimating river loads. Earth Surface Processes and Landforms 1987; 12: 95-104.

Fleischmann N, Staubmann K, Langergraber G. Management of sensible water uses with real-time measurements. Water Sci. Technol. 2002; 46: 33-40.

Forest Technology Systems Ltd. DTS-12 SDI-12 Turbidity Sensor Operating Manual, Victoria, BC, 2007.

Gippel CJ. The use of turbidimeters in suspended sediment research. Hydrobiologia 1989; 176: 465-480.

Gippel CJ. Potential of turbidity monitoring for measuring the transport of suspended solids in streams. Hydrol. Processes 1995; 9: 83-97.

Gray JR, Glysson GD. Proceedings of the Federal Interagency Workshop on Turbidity and other Sediment Surrogates, April 30-May 2, 2002, Reno, Nevada. USGS, 2002, pp. 56.

Grayson RB, Finlayson BL, Gippel CJ, Hart BT. The potential of field turbidity measurements for the computation of total phosphorus and suspended solids loads. J. Environ. Manage. 1996; 47: 257-267.

Grayson RB, Gippel CJ, Finlayson BL, Hart BT. Catchment-wide impacts on water quality: the use of 'snapshot' sampling during stable flow. J. Hydrol. 1997; 199: 121-134.

Gruber G, Winkler S, Pressl A. Continuous monitoring in sewer networks an approach for quantification of pollution loads from CSOs into surface water bodies. Water Sci. Technol. 2005; 52: 215-223.

Harmancioglu NB, Fistikoglu O, Ozkul SD, Singh VP, Alpaslan MN. Water quality monitoring network design. Vol 33. Dordrecht: Kluwer Academic Publishers, 1999.

Heimlich R. Agricultural resources and environmental indicators. US Department of Agriculture Economic Research Service, 2003.

Helsel D, Lee L. Analysis of environmental data with nondetects. Continuing Education Workshop at the Joint Statistical Meetings. American Statistical Association, Seattle, WA, 2006.

Helsel DR. Nondetects and Data Analysis: statistics for censored environmental data. Hoboken: Wiley-Interscience, 2005.

Hem JD. Study and interpretation of the chemical characteristics of natural water. Water Supply Paper 2254. USGS, 1985, pp. 264.

Hooper RP, Goolsby DA, Rickert DA, McKenzie SW. NASQAN, a program to monitor the water quality of the nation's large rivers. USGS, 1997, pp. 6.

Horowitz AJ. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrol. Processes 2003; 17: 3387-3409.

Horsburgh JS, Spackman Jones A, Stevens DK, Tarboton DG, Mesner NO. The Little Bear River Utah, USA: an environmental observatory test bed, 2008. Unpublished manuscript.

Houser JN, Mulholland PJ, Maloney KO. Upland disturbance affects headwater stream nutrients and suspended sediments during baseflow and stormflow. J. Environ. Quality 2006; 35: 352-365.

Johnes PJ. Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. J. Hydrol. 2007; 332: 241-258.

Jordan P, Arnscheidt A, McGrogan H, McCormick S. Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser. Hydrol. Earth System Sciences 2007; 11: 372-381.

Kirchner JW, Feng XH, Neal C, Robson AJ. The fine structure of water-quality dynamics: the (high-frequency) wave of the future. Hydrol. Processes 2004; 18: 1353-1359.

Kronvang B, Bruhn AJ. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrol. Processes 1996; 10: 1483-1501.

Kronvang B, Laubel A, Grant R. Suspended sediment and particulate phosphorus transport and delivery pathways in an arable catchment, Gelbaek Stream, Denmark. Hydrol. Processes 1997; 11: 627-642.

Lewis J. Estimation of suspended sediment flux in streams using continuous turbidity and flow data coupled with laboratory concentrations. Turbidity and Other Surrogates Workshop, Reno, NV, 2002, pp. 1-3.

Lim KJ, Engel BA, Tang ZX, Choi J, Kim KS, Muthukrishnan S, et al. Automated Web based hydrograph analysis tool, WHAT. J. Am. Water Resour. Assoc. 2005; 41: 1407-1416.

McCuen RH. Hydrologic analysis and design. Upper Saddle River, NJ: Prentice Hall, 1998.

Montgomery JL, Harmon T, Kaiser W, Sanderson A, Haas CN, Hooper R, et al. The WATERS network: an integrated environmental observatory network for water research. Environ. Sci. Technol. 2007; 41: 6642-6647.

Mueller DK, Helsel DR. Nutrients in the nation's waters; too much of a good thing? Circular 1136. USGS, 1996, pp. 24.

Neilson BT, Chapra SC. Integration of water quality monitoring and modeling for TMDL development. Water Resour. Impact 2003; 5: 9-11.

Nolan AL, Lawrence GA, Maeder M. Phosphorus speciation in the Williams River, New South Wales: Eutrophication and a chemometric analysis of relationships with other water quality parameters. Marine and Freshwater Res. 1995; 46: 1055-1064.

Oblinger CJ. Triangle Area water supply monitoring project, October 1988 through September 2001, North Carolina- Description of the water-quality network, sampling and analysis methods, and quality-assurance practices. USGS, 2004, pp. 65.

Phillips JM, Webb BW, Walling DE, Leeks GJL. Estimating the suspended sediment loads of rivers in the LOIS study area using infrequent samples. Hydrol. Processes 1999; 13: 1035-1050.

Pressl A, Winkler S, Gruber G. In-line river monitoring - new challenges and opportunities. Water Sci. Technol. 2004; 50: 67-72.

Quilbe R, Rousseau AN, Duchemin M, Poulin A, Gangbazo G, Villeneuve JP. Selecting a calculation method to estimate sediment and nutrient loads in streams: application to the Beaurivage River (Quebec, Canada). J. Hydrol. 2006; 326: 295-310.

Richards RP, Holloway J. Monte-Carlo studies of sampling strategies for estimating tributary loads. Water Resour. Res. 1987; 23: 1939-1948.

Robertson DM, Roerish ED. Influence of various water quality sampling strategies on load estimates for small streams. Water Resour. Res. 1999; 35: 3747-3759.

Ryberg KR. Continuous water-quality monitoring and regression analysis to estimate constituent concentrations and loads in the Red River of the North, Fargo, North Dakota, 2003-05. Scientific Investigations Report. USGS, 2006, pp. 35.

Schaefer DH, Thiros SA, Rosen MR. Ground-water quality in the carbonate-rock aquifer of the Great Basin, Nevada and Utah, 2003. Scientific Investigations Reports. USGS, 2006.

Scholefield D, Le Goff T, Braven J, Ebdon L, Long T, Butler M. Concerted diurnal patterns in riverine nutrient concentrations and physical conditions. Sci. Total Environment 2005; 344: 201-210.

Soil Survey Staff NRCS, USDA. Web Soil Survey, 2008. http://websoilsurbey.nrcs.usda.gov

Spackman Jones A, Stevens DK, Horsburgh JS, Mesner NO. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations, 2008. Unpublished manuscript.

Stubblefield AP, Reuter JE, Dahlgren RA, Goldman CR. Use of turbidometry to characterize suspended sediment and phosphorus fluxes in the Lake Tahoe basin, California, USA. Hydrol. Processes 2007; 21: 281-291.

Tomlinson MS, De Carlo EH. The need for high resolution time series data to characterize Hawaiian streams. J. Am. Water Resour. Assoc. 2003; 39: 113-123.

Uhrich MA, Bragg HM. Monitoring instream turbidity to estimate continuous suspended sediment loads and yields and clay-water volumes in the Upper North Santiam River Basin, Oregon, 1998-2000. USGS, 2003, pp. 44.

US Census Bureau. Downloaded place- and tract-level data from 2000 Census of Population. Bureau of the Census, Department of Commerce, Washington, DC, 2000.

US EPA. Quality Criteria for water 1986. Office of Water Regulations and Standards 440/5-86-001, 1986, pp. 477.

Utah DEQ. Bear River Watershed Management Unit water quality assessment report. Division of Water Quality, 2000a, pp. 41.

Utah DEQ. Little Bear River Watershed TMDL. Division of Water Quality, 2000b.

Utah DEQ. Hyrum Reservoir TMDL Cache County, Utah. Division of Water Quality, 2002.

Utah DEQ. Utah 2006 Integrated Report Volume I- 305(b) Assessment. Division of Water Quality, 2006a, pp. 342.

Utah DEQ. Utah's 2006 Integrated Report Volume II- 303(d) List of Impaired Waters. Division of Water Quality, 2006b, pp. 149.

Utah Water Research Laboratory. Little Bear River WATERS test bed project description, 2007. http://littlebearrriver.usu.edu/description

Vanrolleghem PA, Benedetti L, Meirlaen J. Modelling and real-time control of the integrated urban wastewater system. Environ. Modelling & Software 2005; 20: 427-442.

Vivoni ER, Camilli R. Real-time streaming of environmental field data. Computers & Geosciences 2003; 29: 457-468.

Vivoni ER, Richards KT. Integrated use of GIS-based field sampling and modeling for hydrologic and water quality studies. J. Hydroinformatics 2005; 7: 235-250.

Ward RC, Loftis JC, McBride GB. Design of water quality monitoring systems. New York: Van Nostrand Reinhold, 1990.

WATERS Network. WATERS test bed sites, 2006. http://watersnet.org/wtbs

Webb BW, Phillips JM, Walling DE. A new approach to deriving “best-estimate” chemical fluxes for rivers draining the LOIS study area. Sci. Total Environment 2000; 251: 45-54.

WET Labs Inc. "Cycle" Phosphate sensor, 2006.
http://www.wetlabs.com/products/pub/specshets/cyclessa.pdf

Winkler S, Rieger L, Saracevic E, Pressl A, Gruber G. Application of ion-sensitive sensors in water quality monitoring. Water Sci. Technol. 2004; 50: 105-114.

YSI Inc. 9600 Nitrate monitor, 2006.
https://www.ysi.com/DocumentServer/SocumentServer?docID=EMS_E19

APPENDICES

Appendix A. Plots of Residuals and Statistics of Surrogate Relationships

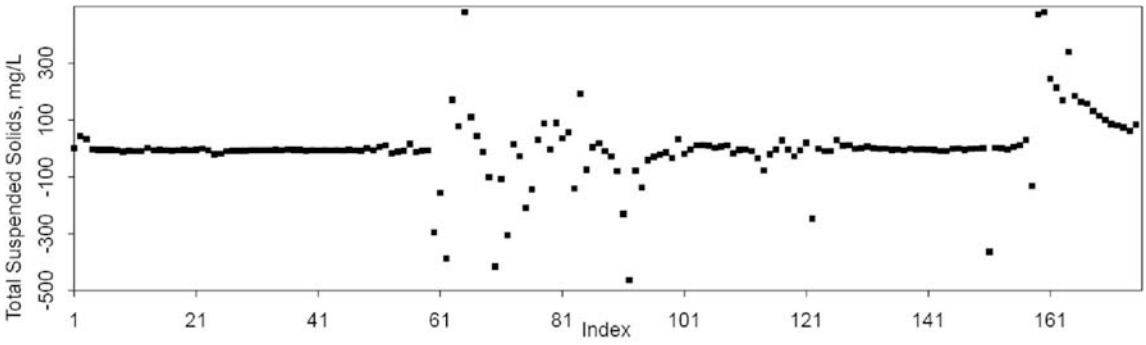

Fig. A-1. Residuals of the TSS model at Paradise.

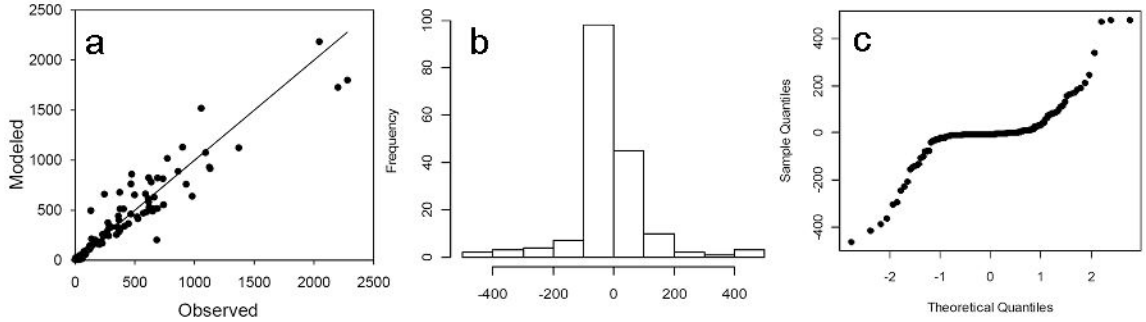

Fig. A-2. Statistical plots for the TSS model at Paradise. Observed versus modeled TSS (a), a histogram (b) and a probability plot (c) of residuals.

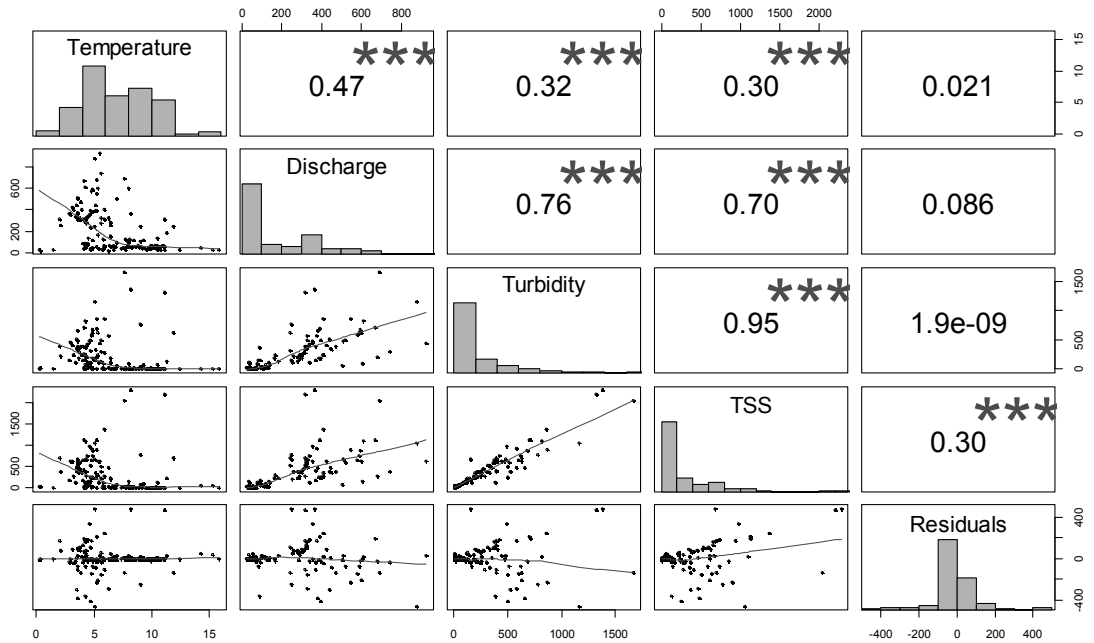

Fig. A-3. Residuals of the Paradise TSS model compared with measured variables. See Fig. 3-2 for interpretation of stars.

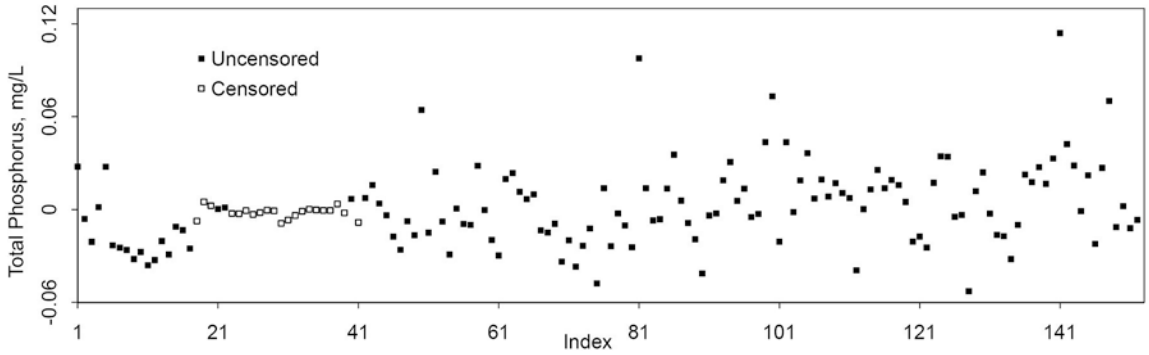

Fig. A-4. Residuals of the TP model at Mendon.

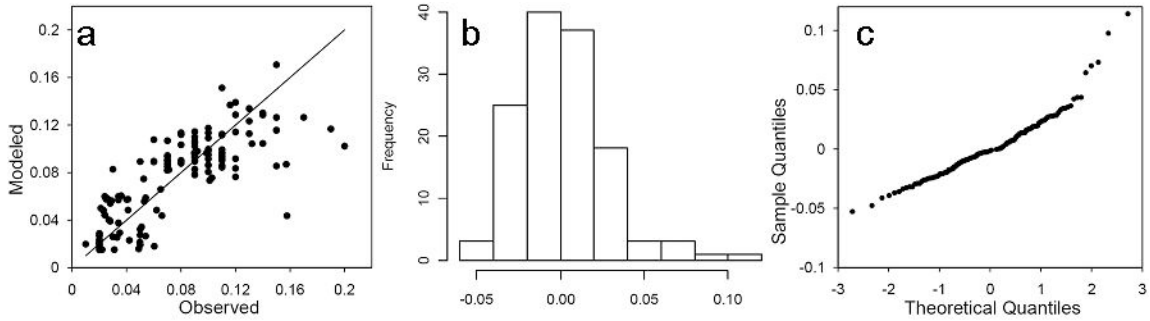

Fig. A-5. Statistical plots for the TP model at Mendon. Observed versus modeled TP (a), a histogram (b) and a probability plot (c) of residuals.

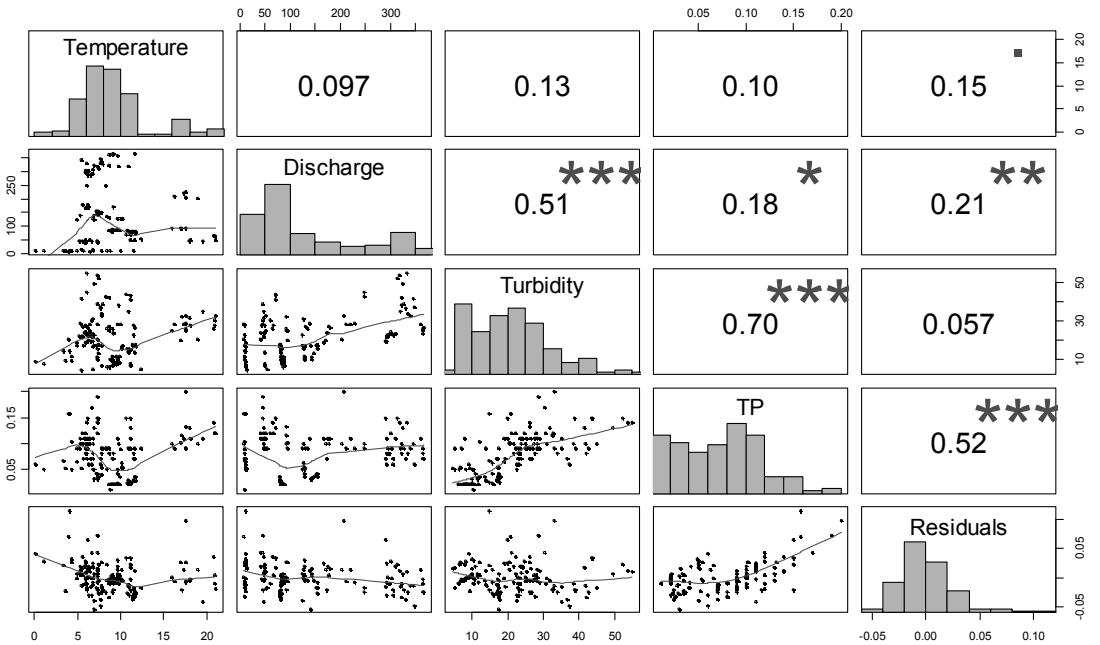

Fig. A-6. Residuals of the Mendon TP model compared with measured variables. See Fig. 3-2 for interpretation of stars.

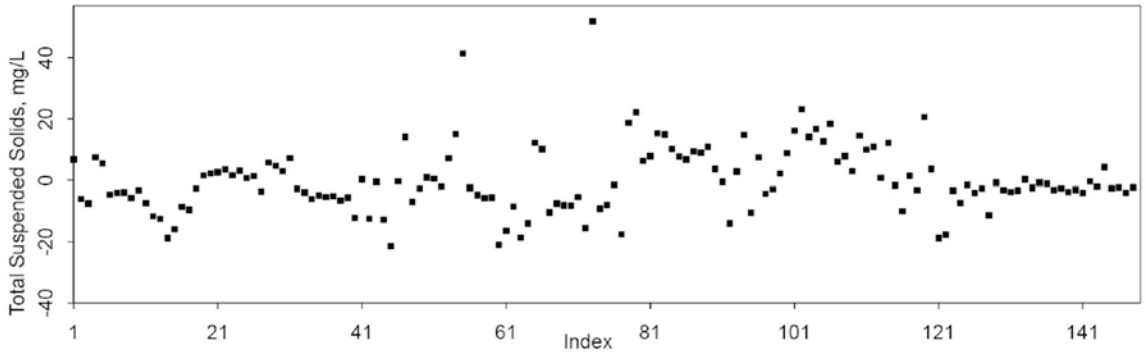

Fig. A-7. Plot of the residuals of the TSS model at Mendon.

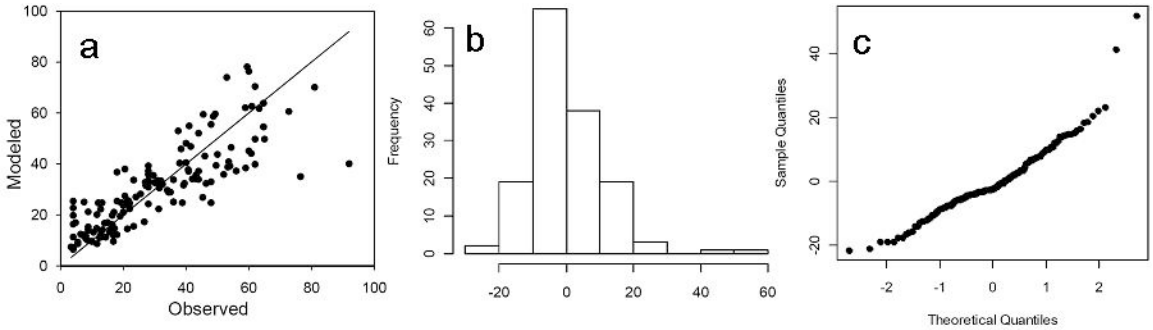

Fig. A-8. Statistical plots for the TSS at Mendon model. Observed versus modeled TSS (a), a histogram (b) and a probability plot (c) of residuals.

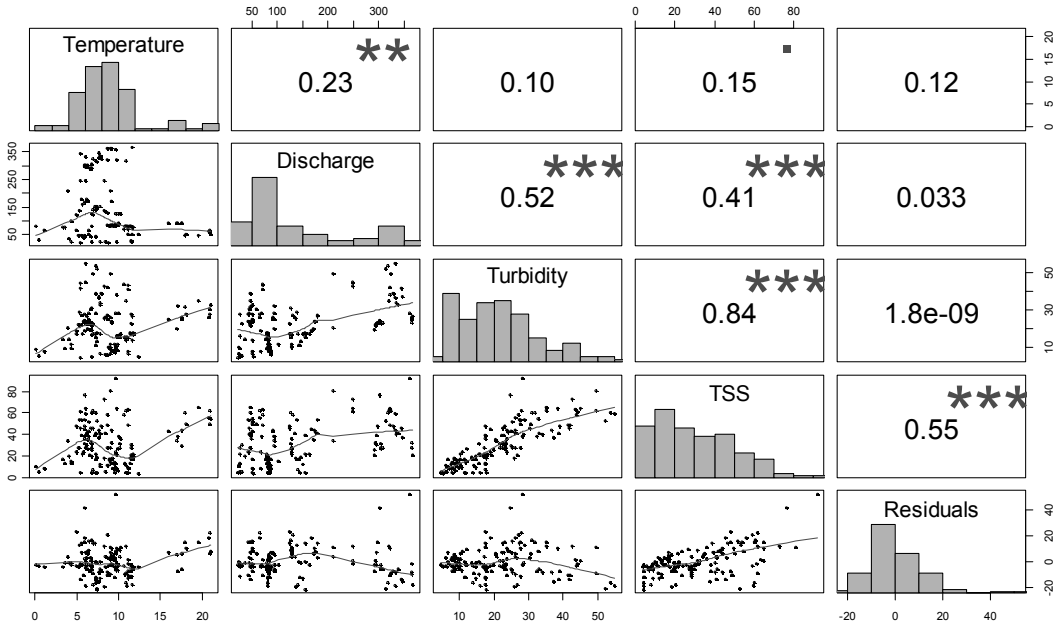

Fig. A-9. Residuals of the Mendon TSS model compared with measured variables. See Fig. 3-2 for interpretation of stars.

Appendix B. Permission Letters from Coauthors

Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322-4110
435 797 0045
October 6, 2008

Jeffrey S. Horsburgh
Utah Water Research Laboratory
Utah State University
Logan, UT 84322-8200

Dear Jeff:

I am in the process of preparing my thesis in the Civil and Environmental Engineering Department at Utah State University. I hope to complete my degree in the fall of 2008.

I am requesting your permission to include the attached papers, of which you are a coauthor, as chapters in my thesis. I will include acknowledgments to your contributions as shown. Please advise me of any changes you require.

Please indicate your approval of this request by signing in the space provided, attaching any other form or instruction necessary to confirm permission. If you have any questions, please contact me.

Thank you,

Amber Spackman Jones

I hereby give permission to Amber Spackman Jones to use and reprint all of the material that I have contributed to Chapters 3 and 4 of her thesis.

Jeffrey S. Horsburgh _____

Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322-4110
435 797 0045
October 6, 2008

Ronald J. Ryel
Department of Wildland Resources
Utah State University
Logan, UT 84322-5230

Dear Ron:

I am in the process of preparing my thesis in the Civil and Environmental Engineering Department at Utah State University. I hope to complete my degree in the fall of 2008.

I am requesting your permission to include the attached paper, of which you are a coauthor, as a chapter in my thesis. I will include acknowledgments to your contributions as shown. Please advise me of any changes you require.

Please indicate your approval of this request by signing in the space provided, attaching any other form or instruction necessary to confirm permission. If you have any questions, please contact me.

Thank you,

Amber Spackman Jones

I hereby give permission to Amber Spackman Jones to use and reprint all of the material that I have contributed to Chapter 4 of her thesis.

Ronald J. Ryel _____