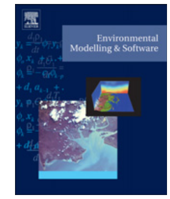
Environmental Modelling & Software 25 (2010) 1031–1044

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft



A sensor network for high frequency estimation of water quality constituent fluxes using surrogates

Jeffery S. Horsburgh [a,*], Amber Spackman Jones [a], David K. Stevens [a], David G. Tarboton [a], Nancy O. Mesner [b]

[a] Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84322-8200, USA
[b] Department of Watershed Sciences, Utah State University, Logan, UT, USA

ARTICLE INFO

Article history:
Received 1 April 2009
Received in revised form
4 October 2009
Accepted 28 October 2009
Available online 14 December 2009

Keywords:
Sensor network
Water quality
Surrogate
Flux
High frequency monitoring
Environmental observatory

ABSTRACT

Characterizing spatial and temporal variability in the fluxes and stores of water and water borne constituents is important in understanding the mechanisms and flow paths that carry constituents to a stream and through a watershed. High frequency data collected at multiple sites can be used to more effectively quantify spatial and temporal variability in water quality constituent fluxes than through the use of low frequency water quality grab sampling. However, for many constituents (e.g., sediment and phosphorus) in-situ sensor technology does not currently exist for making high frequency measurements of constituent concentrations. In this paper we describe how water quality measures such as turbidity or specific conductance, which can be measured in-situ with high frequency, can be used as surrogates for other water quality constituents that cannot economically be measured with high frequency to provide continuous time series of water quality constituent concentrations and fluxes. We describe the observing infrastructure required to make high frequency estimates of water quality constituent fluxes based on surrogate data at multiple sites within a sensor network supporting an environmental observatory. This includes the supporting sensor, communication, data management, and data storage and processing infrastructure. We then provide a case study implementation in the Little Bear River watershed of northern Utah, USA, where a wireless sensor network has been developed for estimating total phosphorus and total suspended solids fluxes using turbidity as a surrogate.

1. Introduction

Characterizing spatial and temporal variability in the fluxes and stores of water and water borne constituents is important in understanding the mechanisms and flow paths that carry constituents to a stream and through a watershed (Montgomery et al., 2007; Wilkinson et al., 2009). Our ability to predict watershed response, which is becoming increasingly important as we work to manage growing pressures on limited water resources, is dependent upon our knowledge of watershed behavior and the interacting processes that drive that response. In some watersheds, the time scale of many important hydrologic and water quality processes is on the order of minutes to hours, not weeks to months (Tomlinson and De Carlo, 2003), and understanding the process linkages between catchment hydrology and stream water chemistry, which is necessary for incorporating these processes into predictive models, requires measurements on a time scale that is consistent with these processes (Kirchner et al., 2004).

Indeed, the need for high frequency monitoring is well recognized (Kirchner et al., 2004; Kirchner, 2006; Hart and Martinez, 2006) and has motivated community initiatives (e.g., http://www.cuahsi.org, http://cleaner.ncsa.uiuc.edu, http://www.watersnet.org/) towards the establishment of large-scale environmental observatories. The goal of these observatory initiatives is to create a network of instrumented sites where data are collected with unprecedented spatial and temporal resolution, aiming at creating greater understanding of the earth's water and related biogeochemical cycles and enabling improved forecasting and management of water processes (Montgomery et al., 2007). Within observatories, environmental sensor networks have been proposed as part of the cyberinfrastructure required to generate data of both high spatial and temporal frequency.

Estimating the flux, or mass flow rate, of a water quality constituent requires estimates of both the constituent concentration and the volumetric flow rate, or discharge of a stream. High frequency monitoring of stream discharge has long been practiced by

* Corresponding author. Tel.: +1 435 797 2946; fax: +1 435 797 3663.
E-mail address: jeff.horsburgh@usu.edu (J.S. Horsburgh).

1364-8152/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2009.10.012

organizations like the United States Geological Survey (USGS) due to the relatively simple methods and technology used to gage discharge and the established need for water quantity measurements in managing water resources (Nolan et al., 2005). However, high frequency discharge monitoring is done at a relatively small number of sites due to the costs associated with continuous monitoring. Traditional water quality monitoring, on the other hand, is generally done at a much greater number of sites, but involves the collection and analysis of grab samples that are usually collected with a frequency too low to accurately characterize the temporal variability in concentrations of water quality constituents (Etchells et al., 2005; Scholefield et al., 2005). Even though both discharge and concentrations are required for estimating constituent fluxes, there is a spatial and temporal disconnect between the traditional methods of monitoring these variables.

High frequency, in-situ monitoring can capture time periods and characterize trends that may be omitted or overlooked by periodic grab sampling (Kirchner et al., 2004; Tomlinson and De Carlo, 2003; Jordan et al., 2007). For many water quality constituents (e.g., sediment and phosphorus), though, sensor technology does not currently exist for making high frequency measurements of concentrations in-situ. For these constituents, it is impossible or impractical to collect samples with high frequency for extended periods due to cost or logistical constraints, leaving us with a paucity of data available for testing models and fostering process understanding (Gascuel-Odoux et al., 2009; May and Sivakumar, 2009). Because of the limitations in existing sensor technology, many studies have examined the use of variables that can be measured in-situ with high frequency as surrogates for other water quality constituents that cannot economically be measured with high frequency.

In this paper we examine the use of surrogates for providing high frequency estimates of water quality constituent fluxes for implementation within sensor networks supporting environmental observatories. We focus on the observing infrastructure and methods required to establish monitoring sites, transmit data, develop site specific surrogate relationships, and the supporting cyberinfrastructure required for developing continuous time series of discharge and concentration for water quality constituents that cannot be measured directly in-situ. We then demonstrate how this infrastructure enables quantification of the spatial and temporal variability in water quality constituent fluxes in ways that cannot be accomplished using low frequency data. Section 2 provides background and discusses the use of surrogate measures for estimating water quality constituent fluxes. In Section 3 we discuss the functional requirements of the required observing infrastructure. Sections 4 and 5 present a specific case of the general problem and describe estimation of total suspended solids (TSS) and total phosphorus (TP) fluxes from continuous water level and turbidity data collected using a wireless sensor network in the Little Bear River of northern Utah, USA. Finally, in Section 6 we summarize our results.

2. Surrogate measures for estimating water quality constituent fluxes

The mass flux of a water quality constituent in a river can be expressed as the product of the constituent concentration and the discharge. In many cases, neither discharge nor concentration can be measured directly in-situ, limiting our ability to estimate fluxes with high frequency. This constraint is primarily due to limitations in existing sensor methods and technology. Measuring stream discharge directly is difficult, and sensors are available for only a small number of water quality constituents. However, surrogates, which can be accurately measured with high frequency in-situ, can enable estimates of both discharge and concentration with high frequency.

Water level, or stage, is a common surrogate that is widely used as an analog for discharge under the premise that discharge in a river increases as the depth of flow increases (McCuen, 2005; Nolan et al., 2005). For water quality constituents, a number of different variables have historically been used as surrogates for concentration. Discharge, for example, has been used as a surrogate to estimate water quality constituent concentrations through rating curves. However, a number of studies have concluded that discharge alone is an unsatisfactory surrogate for constituents such as TSS and TP (Phillips et al., 1999; Robertson and Roerish, 1999; Quilbe et al., 2006; Johnes, 2007). Specific conductance, which is a measure of the ability of water to conduct an electrical current, has been used as an in-situ surrogate for dissolved solids concentrations and for concentrations of ions such as nitrate, sulfate, chloride, and others (e.g., Christensen et al., 2000; Christensen, 2001; Ryberg, 2006). Turbidity, which is an optical measure of the scattering of light passing through water due to colloidal and suspended matter, is also a common surrogate that has been used for total suspended solids, total phosphorus, total nitrogen, and fecal coliform bacteria concentrations (e.g., Christensen et al., 2002; Ryberg, 2006; Stubblefield et al., 2007; Uhrich and Bragg, 2003). The choice of a water quality surrogate depends on the properties of the constituent to be estimated. For example, dissolved constituents are more likely to be better predicted using specific conductance than particulate constituents, which would be better estimated using turbidity.

In concept, most existing commercial sensors work using surrogates. For example, some dissolved oxygen sensors measure the electrical current that is produced as oxygen is reduced at a cathode as more oxygen diffuses through a thin membrane. Since the electrical current is directly proportional to the dissolved oxygen concentration, the sensor converts the current measurement into oxygen concentration units. Similarly, use of a surrogate to estimate discharge or concentration of a water quality constituent relies on there being a strong correlation between the value of the surrogate measurement and the discharge or constituent concentration. The surrogate sensor, coupled with a relationship that converts the sensor's measurements to estimates of discharge or concentration, then, effectively becomes a sensor for the variable of interest when no in-situ sensor exists.

3. Required observing infrastructure for estimating fluxes using surrogates

Robust observing infrastructure is required for making high frequency estimates of water quality constituent fluxes using surrogates due to the large volume of data generated and the need to create continuous records. This is especially true when estimates are needed at many different locations, as will be the case in the proposed environmental observatories (Montgomery et al., 2007; WATERS Network, 2008). Environmental sensor networks are well suited for this task because they enable continuous, high frequency monitoring, they can reduce the logistics and personnel required for grab sampling to be representative (Grayson et al., 1997), and they can help to eliminate errors in transcription and delays in obtaining data (Vivoni and Camilli, 2003).

An environmental sensor network consists of an array of sensor nodes that collect data autonomously and a communications system that allows data collected at each node to reach a computer server (Hart and Martinez, 2006). The required sensor network infrastructure for estimating fluxes from surrogates can be divided into a number of levels, or tiers. The first tier is comprised of sensor and monitoring infrastructure that provides the observational data at locations of interest. The second tier is the communications layer required to support data transmission, monitoring of data collection status, and, in some cases, control of data collection at sensor nodes. The third tier is comprised of data storage, manipulation, and transformation tools that are invoked once sensor data reach a central location. Fig. 1 shows an overall schematic of this three-tiered approach, and, in the following sections, we describe the functional requirements for each tier.

3.1. Tier 1: sensor and monitoring requirements

At each monitoring site, or sensor node, where constituent fluxes are to be estimated, a set of in-situ sensors is needed for providing continuous measurements of the surrogate variables used to estimate both discharge and water quality constituent concentrations. Sensors should be positioned so that their measurements are representative of the flux that is being monitored (e.g., as close to the main flow of a stream as possible), while ensuring that they are hardened against damage from adverse environmental conditions. This can be challenging, especially at sites where there is no permanent structure such as a bridge to mount sensors to. At remote locations with no existing power or communications infrastructure, sensor nodes need to be self powered, capable of unattended data collection, capable of storing data collected between scheduled data downloads, and capable of communicating data and messages to a centralized location. Sensor nodes are most often battery powered, with solar recharge capability and wireless communications capability.

Periodic measurements of discharge and concentration of the water quality constituent of interest are also needed at each sensor node so that surrogate relationships can be derived. The number and frequency of periodic measurements needed depends on: 1) the range of discharge or concentrations experienced in the stream; 2) the nature of the relationship between the surrogate and the variable of interest (e.g., linear versus non-linear); and 3) the desired level of certainty in the resulting surrogate relationship. Discharge measurements are most often made using the area-velocity method described by Buchanan and Somers (1969). For water quality constituents, samples may be collected manually by visiting each site or by using automated samplers to collect samples during storm or snowmelt events. Concentrations are typically determined through laboratory analysis of the samples. Periodic measurements should be made over the range of values to be predicted so that the resulting relationship is representative of the values to be estimated using the relationship.

3.2. Tier 2: communications requirements

The functionality required for sensor network communications infrastructure includes transmission of data from sensor nodes to a centralized location where they can be processed and used and remote monitoring of sensor node status and performance. Additionally, in more sophisticated applications, communication with sensor nodes may be required for remotely adjusting data collection frequency or remotely triggering event based sampling. Although estimation of water quality constituent fluxes from surrogates at a small number of sites does not necessarily require communications infrastructure, scaling to a large number of sites within an environmental observatory will. In many past applications (and even some current ones), a maintenance team visited each monitoring site to manually download data and perform any required site maintenance (Wagner et al., 2006). This approach exposes sensor nodes to potentially long periods of data loss if malfunctions occur between site visits, it precludes any between-visit adjustments to sampling programs, and as the number of sites grows so do the requirements for maintenance personnel. Although the need for sensor node maintenance cannot be completely overcome because sensors must be calibrated and serviced, the ability to remotely monitor sensor node status and retrieve data is invaluable in ensuring the integrity of the continuous data streams.

With recent advances in communications technologies it is becoming easier and more affordable to establish two-way communications with remote sensor nodes (Glasgow et al., 2004).

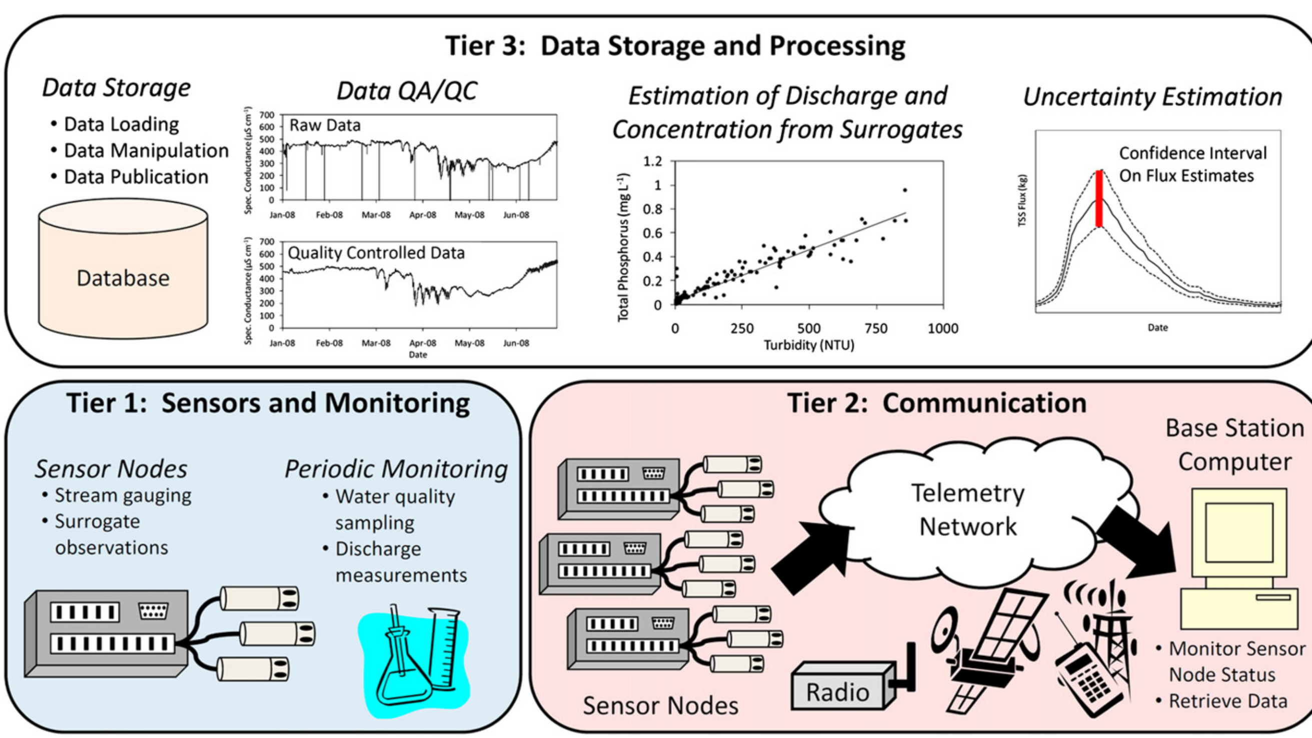


Fig. 1. Overall architecture of the observing infrastructure required to support estimation of water quality constituent fluxes using surrogates.

A number of different technologies are available for building networks to connect sensor nodes (e.g., radio frequency, cellular phone, satellite). The specific technology chosen and the complexity of the resulting data transmission system depends on several factors, including the number of data collection nodes, line-of-sight limitations from the physical terrain in which the sensors are located, the frequency with which data need to be transmitted, and the degree to which sensor nodes need to communicate with each other (e.g., using sensor nodes as repeaters or routers for other sensor nodes). In many cases, communication networks may be made up of a combination of technologies (e.g., radio frequency and cellular phone) as shown in Fig. 2.

3.3. Tier 3: data storage and processing

3.3.1. Data storage

Once data have been transmitted to a central location, a number of manipulation and transformation steps may be required before the data can be used in scientific analyses. The required functionality of data storage infrastructure is to: 1) provide a structure that supports the necessary manipulation, transformation, and subsequent analysis steps; and 2) provide a mechanism for permanent storage of the continuous sensor measurements, results from the periodic samples, the derived flux estimates, and all of the associated metadata necessary for interpreting the data so that they can be made available to users for analyses. This functionality can be facilitated by loading the data into a consistently formatted database.

Databases supporting sensor networks within experimental watersheds have ranged from relatively simple, file and directory based data stores to complex relational databases (e.g., Bosch et al., 2007; Nichols and Anson, 2008). Modern relational database management systems (RDBMS) provide capability for storing, querying, and manipulating data, including a standard query language, whereas file and directory based systems require development of customized software tools for performing these tasks (Connolly and Begg, 2005). Relational databases also facilitate simultaneous access to data by multiple users, which is something that file-based systems have historically done poorly, and can enable publication of the data on the Internet using systems like the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS) (http://his.cuahsi.org). The CUAHSI HIS publishes data stored within a relational database on the Internet using Web services that transmit the data using a standard markup language called WaterML (Zaslavsky et al., 2007; Maidment, 2008; Horsburgh et al., 2009b).

3.3.2. Data quality assurance and quality control

Before sensor data can be used for most applications and analyses they have to be passed through a set of quality assurance and quality control (QA/QC) procedures to ensure that anomalies and spurious data values are removed (Mourad and Bertrand-Krajewski, 2002). In-situ sensors operating in harsh environments occasionally malfunction, some sensors are prone to fouling and drift, and dataloggers and communication systems can corrupt data (Wagner et al., 2006). Uncorrected errors can adversely affect the value of the data for scientific applications, especially if they are to be used by investigators who would have difficulty correcting for these errors because they are not directly familiar with the measurement methods and conditions that may have caused the anomalies (Horsburgh et al., 2009a). QA/QC procedures generally include correction of out of range values, correction for instrument fouling and drift, correction of anomalous values, and correction of any known bias in the sensor data. Wagner et al. (2006) provide guidelines and standard procedures for correcting errors in continuous water quality data streams. Fig. 3 shows examples of raw surrogate data containing errors that must be corrected using QA/QC procedures.

Out of range values (Panel (A) of Fig. 3) beyond the values possible for a variable are generally caused by sensor or equipment malfunctions. Excepting permanent sensor or equipment failures, out of range values are usually short-lived (e.g., one or two simultaneous data values) and can be corrected by interpolation using adjacent data. Known systematic bias in sensor measurements can be corrected by adding an appropriate constant offset to the raw data. Sensor fouling and drift occurs gradually over time, leading to apparent shifts or jumps in the data when sensors are cleaned and

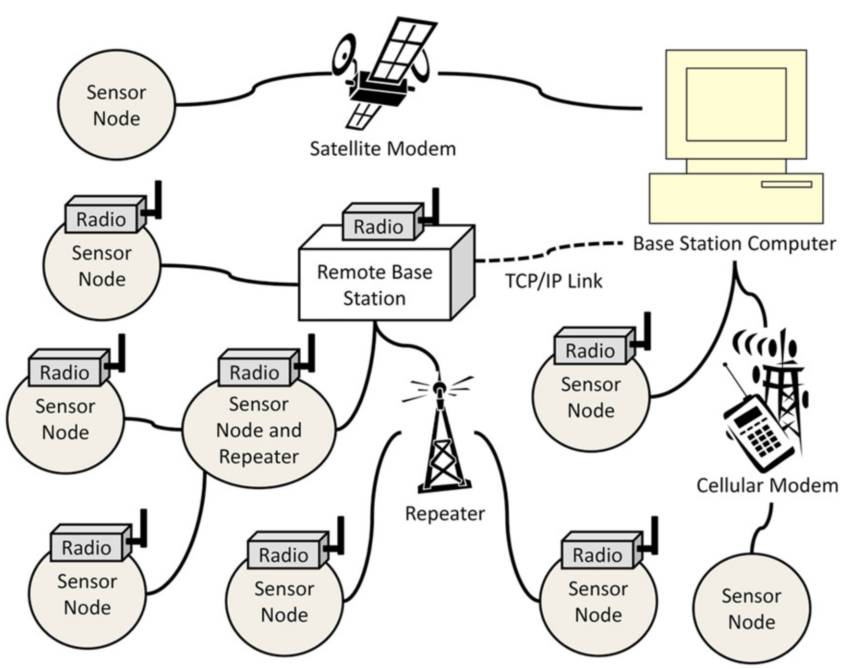


Fig. 2. Common communication pathways in environmental sensor networks.

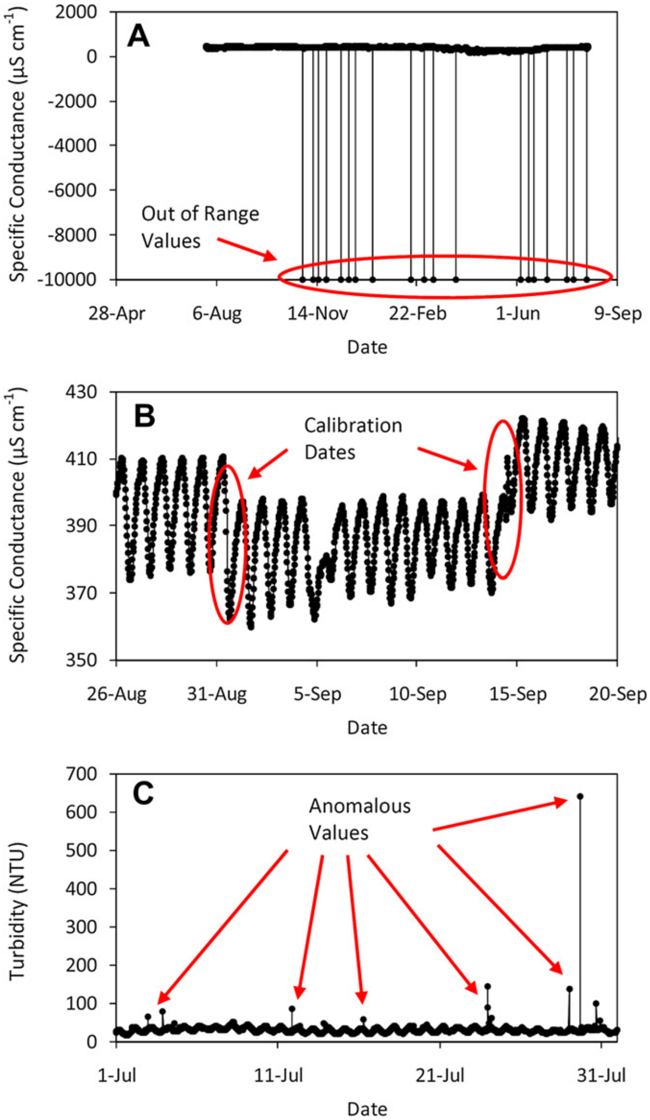


Fig. 3. Errors in observational data that must be corrected using QA/QC procedures. Panel (A) shows out of range values. Panel (B) shows shifts in data caused by sensor drift and calibration. Panel (C) shows anomalous values.

recalibrated (Panel (B) of Fig. 3). The offset, or gap, in the data that occurs when a sensor is recalibrated must be closed, and this is typically done by making an assumption about the nature of the drift (e.g., that it grows linearly over time) and then using that assumption to back-propagate a drift correction. Equation (2) shows an example of a linear drift correction that can be applied to each measured value within a sensor deployment between two calibration dates:

$$V_c = V + \left(V_f - V_s\right)\left(\frac{T_t - T}{T_t}\right) \tag{1}$$

where V_c is the drift corrected data value, V is the original measured data value, V_f is the reading of the sensor immediately before cleaning and calibration at the end of the deployment, V_s is the reading of the sensor immediately after cleaning and calibration, T_t is the total time of the deployment since cleaning and calibration, and T is the time between the end of the deployment and the measured data value.

Anomalies are values that are within the measurement range for a particular variable but significantly different than adjacent data values (Panel (C) of Fig. 3). Anomalies can be artificial (e.g., debris near the face of a turbidity sensor that causes an artificially high reading), but they can also be the result of real transient events (e.g., a brief but intense storm washes sediment into a stream and causes turbidity to rise for a short period) making them much more difficult to correct. Evaluating anomalies can be subjective and requires expertise in both the functioning of the sensor and the processes that drive the sensor response. Comparison with data series from other sites during the same time period can assist in identifying artificial anomalies. Similar to out of range values, short-lived, artificial anomalies can be removed by interpolation using adjacent values.

A number of studies have investigated automated anomaly and error detection in sensor data streams, which is particularly important in real time applications of the data and in detecting instrument malfunctions (Hill et al., 2007; Liu et al., 2007; Mourad and Bertrand-Krajewski, 2002). These methods are generally good at detecting and flagging potentially erroneous sensor values (e.g., out of range values), but because evaluation of anomalies within the measurement range can be subjective, automated procedures may lack the skill to interpret and fix anomalous values. In any case, producing high quality, continuous data streams from raw sensor output can be time and labor intensive, and in many cases involves both automated and manual tools.

3.3.3. Requirements for estimation of discharge and concentration from surrogate measurements

After surrogate data have been corrected using QA/QC procedures, they must then be converted to discharge and concentration using surrogate relationships derived from the periodic measurements of discharge and water quality samples that have been collected. Surrogate relationships can be developed using least squares regression within statistical software. Although statistical software can easily generate regression equations, the appropriateness of the surrogate, other potential explanatory variables, and the resulting regression parameters should be carefully examined. Here we summarize factors that should be considered in developing surrogate relationships. In practice, a more comprehensive text should be consulted (e.g., Helsel and Hirsch, 2002).

A predictive relationship for discharge must be derived for each sensor node using a surrogate such as stage. Although stage–discharge relationships are dependent upon channel geometry and the hydraulic conditions at each site, these relationships typically take the form of a power function as shown in Equation (2) (McCuen, 2005; Nolan et al., 2005):

$$Q = bh^m \tag{2}$$

where Q is discharge ($m^3\ s^{-1}$), h is stage (m), and b and m are constants defining the relationship. Multiple stage–discharge relationships may be required for a single site in cases where one relationship is not appropriate over the entire range of discharges (Nolan et al., 2005). As channel geometry at a site changes over time in response to high flow and sedimentation events, the stage–discharge relationship can change as well. Because of this, stage–discharge relationships must be maintained by continually collecting discharge measurements and refining the relationship as needed.

In some cases, relationships derived between surrogate measurements and water quality constituent concentrations are also site specific (Grayson et al., 1996; Spackman Jones, 2008) and must be developed for each sensor node. For example, turbidity is affected by the scattering properties of suspended particles in water, which are a function of particle size and composition. As a result, a relationship between turbidity and suspended sediment may change with the source of sediment, which varies from site to site (Gippel, 1995; Kronvang and Bruhn, 1996; Tomlinson and De Carlo, 2003; Ryberg, 2006). Like stage–discharge relationships, water quality surrogate relationships can also be dynamic, fluctuating seasonally and over longer time periods as land use and sources of constituent loading change, and should be periodically refined through continued data collection.

Inclusion of multiple surrogates or additional explanatory variables (e.g., season, discharge, temperature, etc.) can be evaluated, although most studies have only included surrogates as predictors if there is a physical basis for the correlation (Christensen, 2001; Rasmussen et al., 2005). Additionally, data transformations (particularly log transformations) are commonly used to achieve constant variance, a linear relationship between independent and dependent variables, or a normal distribution in residuals (Berthouex and Brown, 2002). The need for data transformation varies based on site and constituent. When surrogate relationships have been derived with log transformations, regression-estimated concentrations must be retransformed, introducing retransformation bias. Use of a factor that corrects for this bias can improve predicted concentrations (Helsel and Hirsch, 2002).

There may also be outlier points due to measurement errors or inconsistency between sampled concentration and the water passing in the range of the sensors. Outliers can heavily influence least squares regression and can be considered for removal, but only after they are closely examined. Helsel and Hirsch (2002) describe tests for assessing whether outliers should be omitted from regression analysis as well as regression techniques that are less sensitive to outliers. Table 1 shows water quality surrogate relationships that have been extracted from a number of studies and demonstrates the form that these relationships can take for different constituents.

3.3.4. Uncertainty estimation

In quantifying uncertainty, there are several potential sources of error in the estimated fluxes: 1) measurement error in the surrogate sensor data; 2) measurement error in the periodic observations of discharge and constituent concentrations from which the surrogate relationships are derived; and 3) error in the derived surrogate relationships. Although they do not address uncertainty in in-situ water quality sensor measurements, Harmel et al. (2009) provide an excellent discussion of uncertainty introduced through discharge measurements, water quality sample collection, preservation, storage, and laboratory analysis. Measurement error for in-situ sensors is typically reported by instrument manufacturers, but can also be quantified using multiple instrument tests where resources allow. Measurement error in the periodic observations of discharge and constituent concentrations can be quantified by taking replicate samples and making replicate measurements to derive the components of the measurement error variance (Berthouex and Brown, 2002).

A number of authors have used measures of error in the regressions such as the coefficient of determination (R^2), the root mean square error (RMSE or MSE), and the relative percent difference (RPD) to provide an indication of the uncertainty in the estimates (Christensen et al., 2002; Ryberg, 2006; Stubblefield et al., 2007). Although these statistics are useful for assessing the appropriateness of the surrogate relationships, the uncertainty in the predicted values can be quantified using confidence or prediction intervals (Berthouex and Brown, 2002; Helsel and Hirsch, 2002). Confidence intervals give the range, to a specified degree of confidence, within which the mean value of a response variable is expected to fall. Prediction intervals, on the other hand, give the range, with a specified probability, expected to contain the value of a single new observation of the response variable. Not only does the prediction interval address uncertainty in the derived relationships, but it also incorporates unexplained variance in the response variable (i.e., described above as 2) (Helsel and Hirsch, 2002).

Once the uncertainty in concentration and discharge estimates has been quantified, the uncertainty in the flux can be estimated using one of several error propagation techniques such as first order error analysis, Monte Carlo simulation, or bootstrapping. First order error analysis evaluates the variance in the estimation of constituent flux given the variances of concentration and the discharge, as given by Equation (2):

$$Var(W) = \left(\frac{\partial W}{\partial C}\right)^2_{\bar{C}} Var(C) + \left(\frac{\partial W}{\partial Q}\right)^2_{\bar{Q}} Var(Q) + 2\left(\frac{\partial W}{\partial C}\right)_{\bar{C}}\left(\frac{\partial W}{\partial Q}\right)_{\bar{Q}} Cov(C,Q) \quad (3)$$

where W is the flux, or mass loading, of the water quality constituent, C is concentration, and Q is discharge. The covariance term can be omitted if concentration is independent of discharge; however, the covariance may be negative and reduce the overall variance in the load (Berthouex and Brown, 2002). First order error analysis is appropriate where the surrogate relationships are linear. For non-linear equations or for periods that the stage–discharge relationship is non-linear, simulation techniques are more appropriate.

4. A case study for estimating total phosphorus and total suspended solids fluxes: the Little Bear River sensor network

A wireless sensor network has been established in the Little Bear River of northern Utah, USA that demonstrates a specific case of the general problem of estimating water quality constituent fluxes from surrogates. Using the Little Bear River sensor network, high frequency TSS and TP fluxes are estimated from in-situ turbidity and stage measurements at a number of sensor nodes. Each of the components of the general architecture described above has been applied in the Little Bear River, and here we describe their specific implementations within the Little Bear River sensor network.

4.1. Sensors and monitoring

Seven sensor node locations were selected to characterize the major hydrologic conditions in the Little Bear River watershed and to represent the range of land use conditions, with preference given to locations that would provide the most information given our limited resources. In addition, site selection was dependent on the presence of a bridge or other permanent structure to which the sensors could be mounted, the ability to obtain permission to access the site, the ability to establish a stream cross section suitable for development of a stage–discharge relationship, and the ability to establish communications with the site to retrieve the data. Fig. 4 shows the locations of the sensor nodes within the Little Bear River watershed.

Each sensor node consists of in-situ stage and turbidity sensors connected to a datalogger. The dataloggers were programmed to collect data every 30 min and store it in the datalogger's onboard memory. The dataloggers use the SDI-12 communication protocol to communicate with the each of the sensors. Sensors were installed as close to the main flow of the river as possible and were enclosed within PVC pipe housings to protect them from debris and vandalism (Fig. 5). The PVC sensor housings were fitted with metal pump screens into which the sensors extend to ensure adequate water flow-through and to protect the sample space around each of the sensors. Sensors are removed and cleaned in the field at least once every two weeks.

Table 1
Examples of surrogate relationships for water quality constituents.

Constituent	Surrogates	Regression equation	Source
Alkalinity (ALK)	Discharge (Q), Water Temperature (WT)	$\log ALK = 0.000368Q - 0.000148WT^2 + 2.36$	Christensen (2001)
	Specific Conductance (SC), Discharge (Q)	$ALK = 0.165SC - 54.3\log Q + 261$	Ryberg (2006)
	Specific Conductance (SC)	$\log ALK = 0.516 \log SC + 0.746$	Rasmussen et al. (2005)
Dissolved Solids (DS)	Specific Conductance (SC)	$DS = 0.549SC + 14.3$	Christensen (2001)
	Specific Conductance (SC)	$DS = 0.689SC - 52$	Ryberg (2006)
	Specific Conductance (SC)	$\log DS = 0.966 \log SC - 0.115$	Rasmussen et al. (2005)
Suspended Solids (SS)	Turbidity (TURB)	$\log SS = 0.818 \log TURB + 0.348$	Christensen (2001)
	Discharge (Q), Turbidity (TURB)	$\log SS = 0.213 \log Q + 0.814 \log TURB - 0.092$	Ryberg (2006)
	Turbidity (TURB)	$SS = 1.45{*}TURB^{1.08}{*}1.13$	Uhrich and Bragg (2003)
	Turbidity (TURB)	$\ln SS = 1.04 \ln TURB - 0.535 + 0.326$	Tomlinson and De Carlo (2003)
	Turbidity (TURB)	$SS = 3.29TURB - 6.54$	Stubblefield et al. (2007)
	Turbidity (TURB)	$SS = -0.76 + 0.92TURB$	Grayson et al. (1996)
	Turbidity (TURB)	$TURB = SS^{0.71}$	Kronvang and Bruhn (1996)
Total Nitrogen (TN)	Turbidity (TURB), Water Temperature (WT), Specific Conductance (SC)	$TN = 0.00317TURB + 0.0234WT - 0.0000655SC + 0.469$	Christensen (2001)[a]
	Turbidity (TURB), Discharge (Q)	$TN = 0.0042TURB - 0.000089Q + 0.494$	Christensen et al. (2002)
	Discharge (Q), Day of Year (D)	$TN = 0.422 \log Q + 0.699 \cos(2\pi D/365) - 0.318\sin(2\pi D/365) + 0.4\cos(4\pi D/365) - 0.202 \sin(4\pi D/365) + 0.03$	Ryberg (2006)
	Discharge (Q), Turbidity (TURB)	$\log TN = 0.111 \log Q + 0.0004TURB - 0.0585$	Rasmussen et al. (2008)
Total Phosphorus (TP)	Turbidity (TURB), Specific Conductance (SC), Water Temperature (WT)	$TP = 0.00103TURB - 0.227 \log SC + 0.0057WT + 0.776$	Christensen (2001)
	Turbidity (TURB)	$TP = 0.000606TURB + 0.186$	Christensen et al. (2002)
	Discharge (Q), Turbidity (TURB), Day of Year (D)	$TP = 0.111\log Q + 0.353\log TURB + 0.056\cos(2\pi D/365) - 0.047\sin(2\pi D/365) - 0.734$	Ryberg (2006)
	Turbidity (TURB)	$TP = 26.7 + 1.58TURB$	Grayson et al. (1996)
	Turbidity (TURB)	$TP = 0.0012TURB + 0.152$	Rasmussen et al. (2008)
Fecal Coliform (FC)	Water Temperature (WT), Turbidity (TURB)	$\log FC = -3.4 \log WT + 0.432 \log TURB + 6.53$	Christensen (2001)
	Turbidity (TURB), Discharge (Q), Specific Conductance (SC), Day of Year (D)	$\log FC = -0.527 \sin(4\pi D/365) - 0.82 \cos(4\pi D/365) + 0.0113TURB + 2.2\log Q + 0.00045SC - 3.71$	Christensen et al. (2002)
	Turbidity (TURB)	$\log FC = 1.641 \log TURB - 0.121$	Rasmussen et al. (2008)
Sodium (NA)	Specific Conductance (SC), Discharge (Q)	$NA = 0.203SC + 0.0938Q - 117$	Christensen (2001)
	Specific Conductance (SC)	$\log NA = 1.46 \log SC - 2.39$	Rasmussen et al. (2005)
Chloride (CL)	Specific Conductance (SC), Discharge (Q)	$CL = 0.319SC + 0.113Q - 172$	Christensen (2001)
	Specific Conductance (SC), Discharge (Q)	$CL = -9.55 \log Q + 0.011SC + 38.8$	Ryberg (2006)
	Specific Conductance (SC)	$\log CL = 1.74 \log SC - 3.14$	Rasmussen et al. (2005)
Fluoride (F)	Specific Conductance (SC), Discharge (Q)	$\log F = -0.000255Q + 0.162 \log SC - 0.892$	Christensen (2001)
	Specific Conductance (SC)	$\log F = 0.217 \log SC - 1.1$	Rasmussen et al. (2005)
Sulfate (SO)	Specific Conductance (SC)	$SO = 0.0268SC + 13.17$	Christensen (2001)
	Specific Conductance (SC), Discharge (Q)	$\log SO = 0.128 \log Q + 0.011SC + 38.8$	Ryberg (2006)
	Specific Conductance (SC)	$\log SO = 1.12 \log SC - 1.28$	Rasmussen et al. (2005)

[a] The equation reported is for total organic nitrogen.

Fig. 4. Locations of sensor nodes in the Little Bear River watershed.

At sensor node locations, grab samples for TP and TSS are collected weekly during the spring snowmelt season (March through June) and every other week during the rest of the year to provide the data necessary for deriving surrogate relationships. Additionally, storm event and spring snowmelt event samples have been collected using portable automated samplers to ensure that expected periods of high flux are characterized. Samplers are deployed either when precipitation is expected or when a significant snowmelt event is expected. Manual discharge measurements are made seasonally to ensure that a wide range of flows is captured at each location.

4.2. Communications

Surrogate data from each sensor node are transmitted in near real time to the Utah Water Research Laboratory (UWRL) via a communications network. The network uses 900 MHz spread-spectrum radios for transmitting data from sensor nodes to one of two remote base stations located at public schools within the watershed. Because the distances between sensor nodes are relatively large (up to 7 km) and the watershed has high relief (Fig. 4), two radio repeaters were installed to extend the reach of the network. From the remote base stations, the data are transmitted using Ethernet TCP/IP links (established using Campbell Scientific Network Link Interfaces) to a server at the UWRL.

Communications within the network are managed using Campbell Scientific's Loggernet software (http://www.campbellsci.com). Loggernet enables us to monitor sensor node status in real time, regularly retrieve data from each of the sensor nodes, and send new programs or instructions to each of the sensor nodes from a server located at the UWRL. This communication system was chosen because it uses established commercial technology, eliminates monthly service costs, has relatively low power requirements, and provides us with flexibility for accepting new sites onto the existing network. The LoggerNet server is programmed to connect hourly to each sensor node and download the most recent data to delimited text files.

4.3. Data storage and processing

At the UWRL, the water level and turbidity data and supporting metadata are automatically loaded from the datalogger files into a relational database that implements the CUAHSI HIS Observations

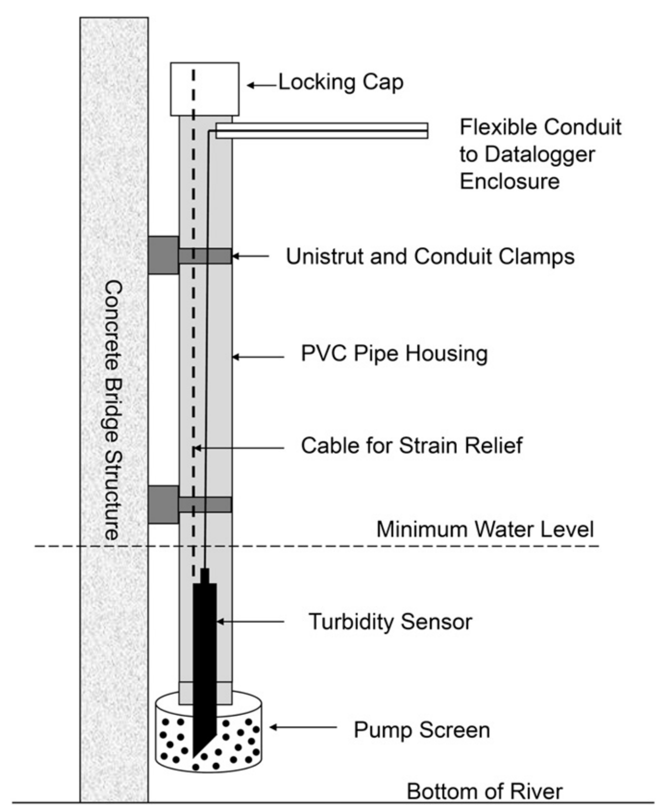


Fig. 5. Schematic of a typical sensor deployment.

Data Model (ODM) (Horsburgh et al., 2008) using the Streaming Data Loader software application, which is also part of the CUAHSI HIS. The Streaming Data Loader maps each of the datalogger files to the ODM schema to ensure that appropriate metadata are associated with the data values and then automatically loads the sensor data into the database. Execution of the Streaming Data Loader is scheduled so that the most recent values are always available in the database.

Once they have been loaded into the database, the surrogate measurements are corrected by a technician using a combination of graphical techniques and the quality control procedures described above. New values are processed approximately once per month. Many of the QA/QC techniques discussed above have been implemented within a software application called ODM Tools (also part of the CUAHSI HIS), which provides the technicians with a graphical user interface for performing quality control of data. ODM Tools provides functionality for removing obvious errors or out of range values, sensor malfunctions, and instrument drift. All corrections and edits are performed on a copy of the raw data to ensure that the original data are preserved. ODM is capable of storing multiple copies, or versions, of each time series, with each version identified by the level of quality control to which it has been subjected. ODM also preserves the provenance of the data by storing the linkage between raw and corrected data values.

Least squares regression was used to develop stage–discharge, turbidity-TP, and turbidity-TSS relationships at each site. Because many of the TP samples collected at each site had concentrations below the detection limit of the method used by the analytical laboratory, regression with maximum likelihood estimation (MLE) was performed using techniques described by Helsel (2005) to account for below detection limit observations. Besides turbidity, additional explanatory variables (e.g., discharge, temperature, hour of day) were examined for significance in the surrogate relationships. Variables representing the hydrologic conditions (i.e., the occurrence of spring snowmelt or a storm event) at the time of sample collection were also explored. The model equations that were ultimately selected provided the minimum root mean square error values, and all of the explanatory variables had p-values within the 95% significance level.

The corrected surrogate data from each sensor node were converted to time series of discharge and concentration using the derived surrogate relationships. The derived time series were stored within the database so that they were available for analyses and so they could be manipulated using the query tools available in the database management system. Finally, TSS and TP fluxes were examined by multiplying the discharge data series by the concentration data series to create time series of TSS and TP fluxes. All of the data collected in the Little Bear River, including the continuous measurements of water level and turbidity, the quality controlled versions of these, the periodic water quality grab sampling results, and the derived datasets (e.g., continuous discharge from water level and TSS and TP from turbidity) have been published using the CUAHSI HIS data publication system (Horsburgh et al., 2009b) and are available via http://littlebearriver.usu.edu.

4.4. Results and discussion

4.4.1. Spatial variability in Little Bear River TSS fluxes

Fig. 6 shows the total annual TSS fluxes at 5 sensor nodes in the Little Bear River watershed for water year 2008. Annual fluxes tend to increase in a downstream direction until the river reaches Hyrum Reservoir, where much of the sediment carried by the river is trapped. This is reflected by a relatively low flux at the node immediately below the reservoir (Node 6 at Wellsville), although this is also due to diversions of water from the reservoir outflow that result in reduced discharge at that node. At the most downstream

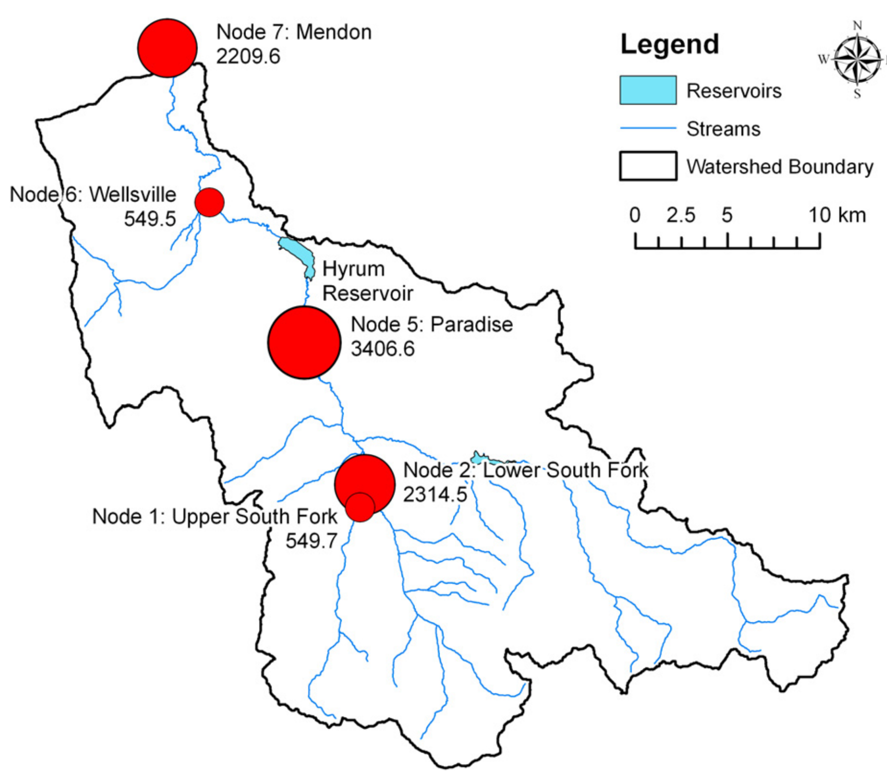


Fig. 6. Spatial distribution of total suspended solids fluxes in the Little Bear River for 2008. The areas of the node markers are proportional to the total suspended solids fluxes, which are expressed in metric tons.

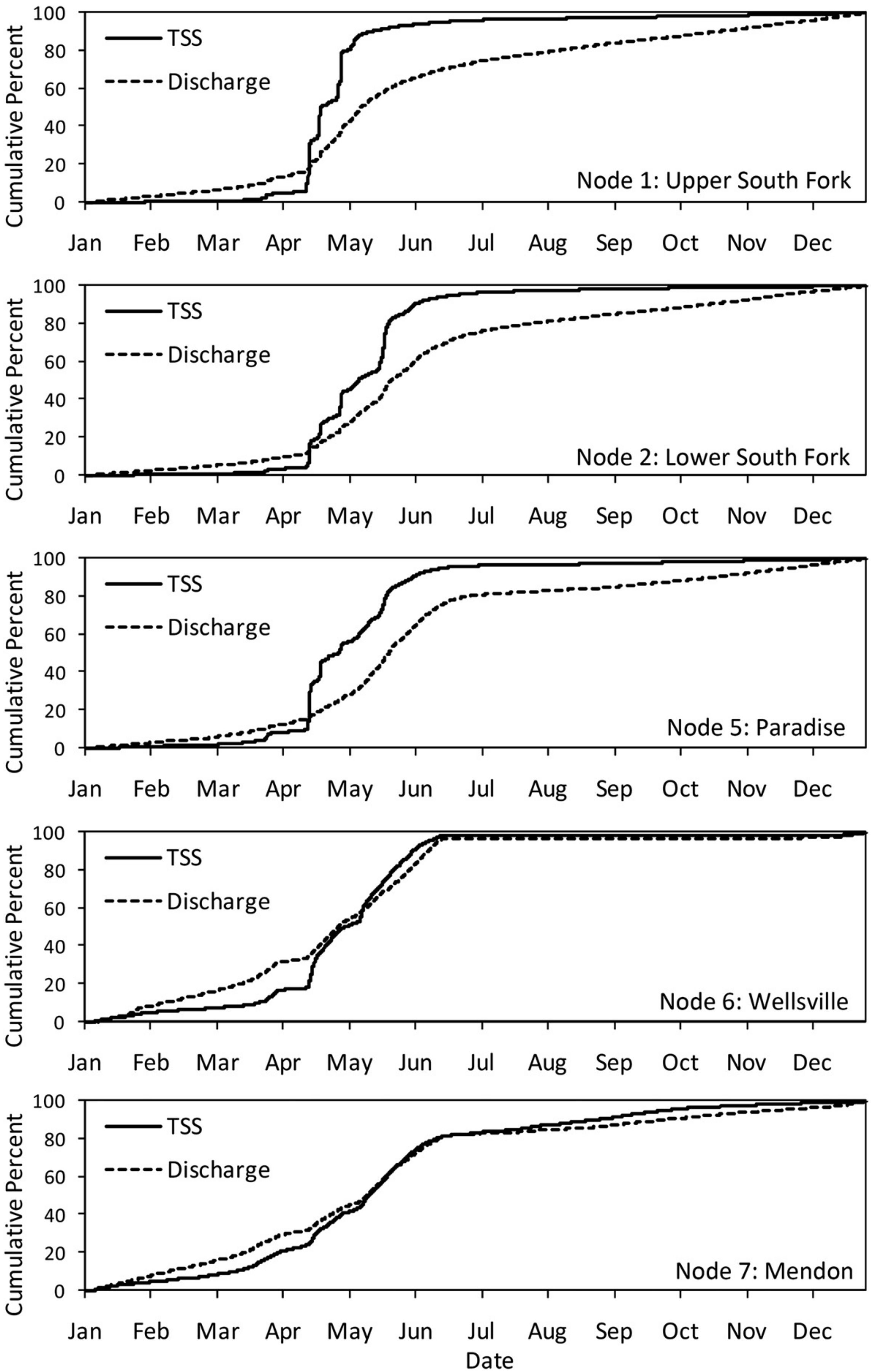


Fig. 7. Timing of discharge and total suspended solids fluxes for several sensor nodes in the Little Bear River during 2008.

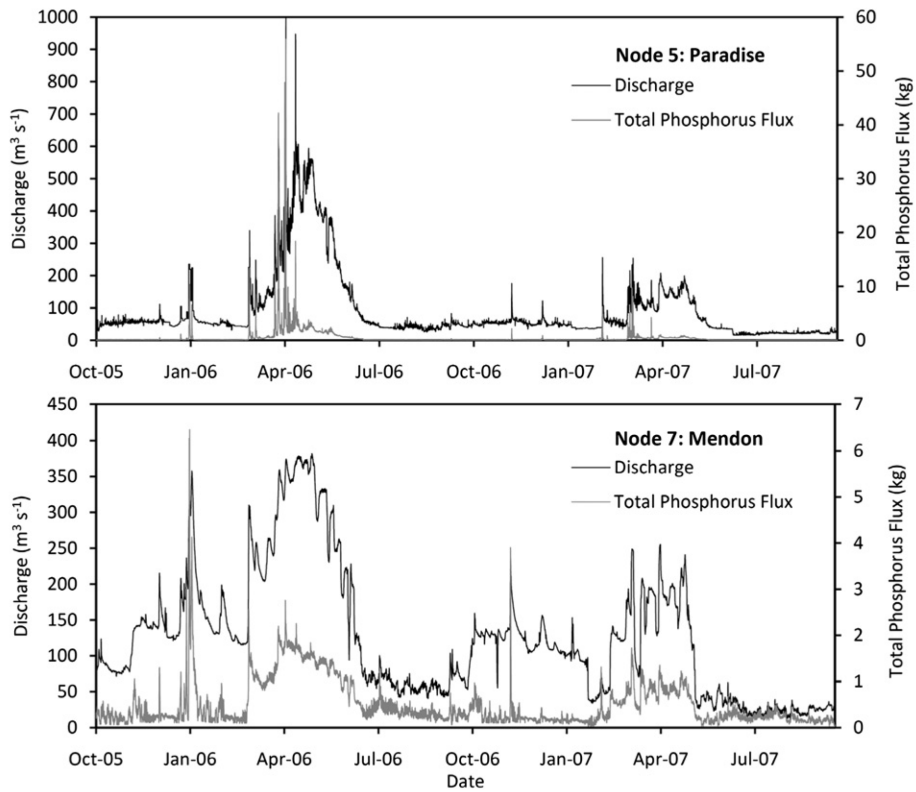


Fig. 8. Discharge and 30-min total phosphorus fluxes for water years 2006 and 2007 at the Mendon and Paradise sensor nodes.

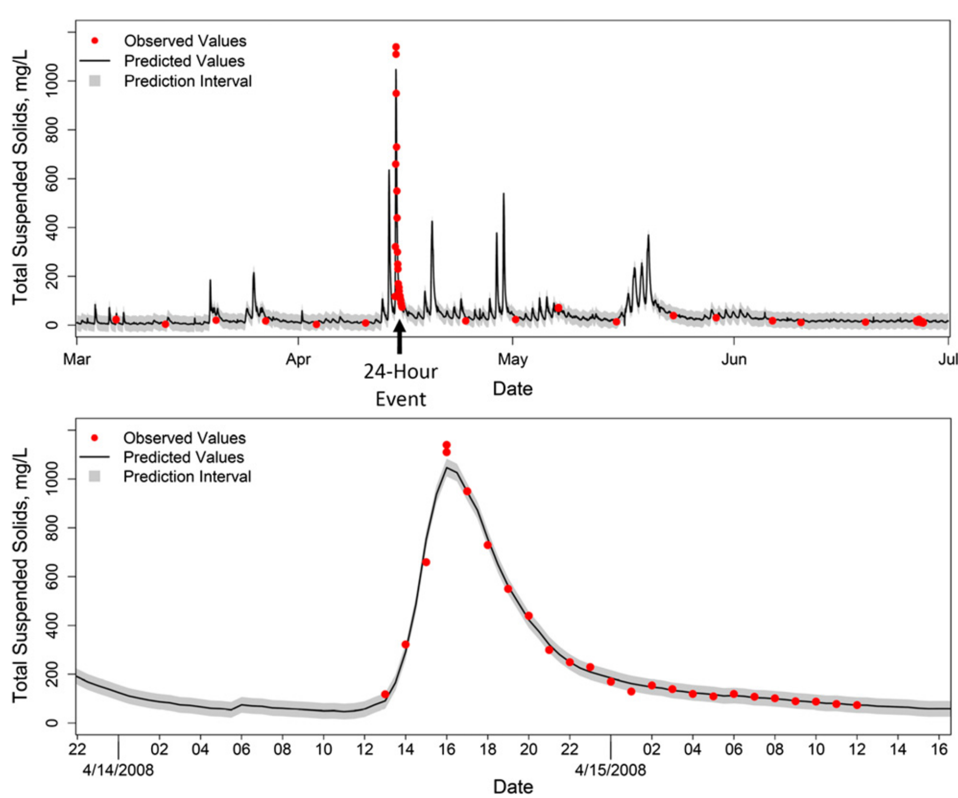


Fig. 9. Total suspended solids concentrations predicted from turbidity for the lower South Fork Little Bear River sensor node (Node 5) during the spring of 2008. The top panel shows weekly TSS samples along with a 24-h snowmelt sampling event on April 14–15. The bottom panel shows a zoomed view of the 24-h sampling event. The grey shaded area shows the 95% prediction intervals for the estimated TSS concentrations.

node (Node 7 at Mendon), the annual flux is again relatively high due to sediment laden agricultural return flows that heavily influence the river.

Fig. 7 shows the cumulative percentage of discharge and TSS flux based on high frequency flux estimates for the year 2008. Five locations in the Little Bear River watershed are shown to illustrate the timing of discharge and TSS flux at different nodes. Nodes in the upper watershed above Hyrum Reservoir (Node 1 Upper South Fork, Node 2 Lower South Fork, and Node 5 at Paradise) show steep flux curves during the spring. At these nodes, the flux curve is steeper than the discharge curve indicating that the timing of the TSS flux is not simultaneous with discharge and is weighted towards the beginning of the spring snowmelt period. The Wellsville node (Node 6) also exhibits a relatively steep flux curve during the spring, but TSS flux and discharge are more similarly timed. Discharge and TSS flux at the Mendon node (Node 7) are similarly timed, but they show a much more gradual slope throughout the year.

The spatial variability and timing of the total annual flux reveal important information about the flow pathways and processes that carry water quality constituents to the stream and through a watershed. At nodes above Hyrum Reservoir, TSS flux is primarily driven by snowmelt, which is reflected in the steep slope of the cumulative flux plots during a relatively short period during the spring when the snow is melting. Much of the TSS flux occurs near the beginning of the snowmelt period, which is likely due to the fact that snow at lower elevations close to the stream channels melts first, carrying TSS to the stream through surface pathways. As snowmelt moves further away from active streams, the water contributing to stream discharge is less likely to carry TSS. Additionally, as discharge and stream velocities increase with snowmelt, sediment stored in the stream channel is transported until the storage of TSS within the channel is exhausted later in the snowmelt period.

Hyrum Reservoir (Fig. 4) serves as a reset point for water quality and effectively divides the Little Bear River watershed in two. TSS flux at Wellsville (Node 6) is driven by spills from Hyrum Reservoir that occur for a short period after it fills in the spring. Consequently, approximately 90% of the annual TSS flux at Wellsville occurs between the months of March and May. After the end of May, nearly all of the discharge that would normally be in the river at Wellsville is either diverted for irrigation (until the end of the irrigation season) or is stored in Hyrum Reservoir, which is why very little discharge or TSS flux occurs after the beginning of June. At Mendon (Node 7), discharge and TSS flux are driven by springtime spills from Hyrum Reservoir and by sediment enriched agricultural return flows during the irrigation season (April–September). Additionally, the river channel between Wellsville and Mendon cuts through fine soil material that contributes to more constant streambed and bank erosional processes that are not as prominent in the upper watershed.

4.4.2. Temporal variability in Little Bear River phosphorus fluxes

Time series of discharge and TP fluxes for the 2006 and 2007 water years at the Paradise (Node 5) and Mendon (Node 7) sensor nodes are shown in Fig. 8. Fig. 8 illustrates the large interannual variability in both discharge and constituent loading that can occur within the Little Bear River. This is particularly apparent in the data for the Paradise sensor node, which is located above Hyrum reservoir and is much more susceptible to variability in natural flows (2007 was a low flow year when compared to 2006). Flows at the Mendon sensor node, which is located below Hyrum reservoir, are increased by agricultural return flows throughout much of the summer, significantly altering the natural flow and TP flux regime. Fig. 8 underscores the importance of monitoring over long periods to better quantify the range of hydrologic conditions that can occur.

4.4.3. Measurement error and uncertainty

Measurement error and uncertainty in the derived surrogate relationships can have a substantial effect on the uncertainty of flux estimates made using surrogates. This should certainly be taken into account when interpreting fluxes derived from surrogates. Fig. 9 shows TSS concentrations predicted from turbidity for the lower South Fork Little Bear River sensor node (Node 5) during the spring of 2008. Also shown are 95% prediction intervals for the TSS estimates, and observed TSS concentrations. The top panel shows weekly TSS samples along with a 24-h snowmelt sampling event on April 14–15 (one sample per hour), and the bottom panel shows a zoomed view of the 24-h snowmelt event. In these plots, the TSS concentrations predicted using the surrogate relationship agree well with the observed concentrations. The prediction intervals show the uncertainty in the predicted concentrations resulting from uncertainty in the surrogate relationship.

Fig. 9 also shows the dynamic nature of turbidity and TSS at this site and demonstrates how weekly TSS observations do not capture this variability. While the hourly samples in the 24-h snowmelt event do well at characterizing one of the days where TSS concentrations ranged from approximately 150–1200 mg/L, sampling every hour to characterize all of the peaks would be cost prohibitive. Despite the uncertainty in the surrogate relationships, concentrations and fluxes estimated using continuous surrogate data are preferable to the alternative of estimates based on a handful of samples because the continuous data capture the dynamics of the system at a time scale that is consistent with the processes that are occurring (in this case daily cycles in spring snowmelt).

5. Conclusions

In this paper, we have presented the observing infrastructure and methods needed for making long-term, high frequency estimates of water quality constituent fluxes from surrogates. Our examples from the Little Bear River show how high frequency data that are consistent with the spatial and temporal scale of processes that control variability in the fluxes and stores of water and water borne constituents can assist us in better understanding the mechanisms and flow paths that carry constituents through watersheds. However, until sensor technology advances to the point where affordable and reliable in-situ sensors are available for all of the water quality constituents in which we are interested, high frequency estimation of constituent fluxes in streams and rivers will likely rely on existing surrogate sensors.

The Little Bear River sensor network case study demonstrates a specific implementation of the sensor and monitoring, communications, and data storage and processing infrastructure required for creating a network of flux monitoring sites. As sensor networks continue to be implemented in support of scientific research within environmental observatories and at other data-intensive research sites, the need will grow for robust and automated infrastructure for collecting and storing the large data volumes, monitoring data collection status, correcting and processing the data, and making the data available to analysts. The innovative combination of methods and tools that we have described, including the components of the CUAHSI HIS that we adopted to support our work, certainly advances capabilities for implementation at other sites.

The Little Bear River sensor network shows how the common spatial and temporal disconnect between traditional methods of monitoring discharge and water quality constituent concentrations can be overcome. The specific results from the Little Bear River that we have presented are examples of the types of analyses that are enabled by implementing the observing infrastructure that we have described and demonstrate the value of high frequency flux estimates in furthering our understanding of water quality constituent fluxes and important flow pathways that carry them. Future refinements of the surrogate methods that we have presented are needed to ensure that the sampling protocols are efficient – i.e., minimizing costs while achieving acceptable accuracy in the resulting flux estimates. This will involve better methods for deciding how many grab samples are needed to establish and maintain surrogate relationships and using adaptive monitoring to decide when to collect those samples so that they contain the most information.

Acknowledgments

This work was supported by the National Science Foundation grant CBET 0610075 and by the United States Department of Agriculture's Conservation Effects Assessment Program (CEAP) grant UTAW-2004-05671. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the United States Department of Agriculture. Mention of trade names or commercial products within this paper is for identification purposes only and does not constitute endorsement or recommendation for use by the National Science Foundation, the United States Department of Agriculture, or Utah State University.

References

Berthouex, P.M., Brown, L.C., 2002. Statistics for Environmental Engineers, second ed. Lewis Publishers, New York, 489 pp.

Bosch, D.D., Sheridan, J.M., Lowrance, R.R., Hubbard, R.K., Strickland, T.C., Feyereisen, G.W., Sullivan, D.G., 2007. Little river experimental watershed database. Water Resources Research 43, W09470. doi:10.1029/2006WR005844.

Buchanan, T.J., Somers, W.P., 1969. Discharge measurements at gaging stations, Book 3, Chapter A8, applications of hydraulics. In: Techniques of Water-resources Investigations of the United States Geological Survey. United States Department of the Interior, Washington, DC. Available at: http://pubs.usgs.gov/twri/twri3a8/html/pdf.html.

Christensen, V.G., 2001. 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, USGS Water Resources Investigations Report 01-4248, 28 pp. Available at: http://ks.water.usgs.gov/pubs/reports/wrir.01-4248.pdf.

Christensen, V.G., Jian, X., Ziegler, A.C., 2000. 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, USGS Water Resources Investigations Report 00-4126, 36 pp. Available at: http://ks.water.usgs.gov/pubs/reports/wrir.00-4126.html.

Christensen, V.G., Rasmussen, P.P., Ziegler, A.C., 2002. Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams. Water Science and Technology 45 (9), 205–211.

Connolly, T., Begg, C., 2005. Database Systems: a Practical Approach to Design, Implementation, and Management, fourth ed. Addison-Wesley, Harlow, UK.

Etchells, T., Tan, K.S., Fox, D., 2005. Quantifying the uncertainty of nutrient load estimates in the Shepparton irrigation region. In: Proceedings MODSIM05 International Congress on Modelling and Simulation, Advances and Applications for Management and Decision Making. Available at: http://www.mssanz.org.au/modsim05/papers/etchells_2.pdf.

Gascuel-Odoux, C., Aurousseau, P., Cordier, M., Durand, P., Garcia, F., Masson, V., Salmon-Monviola, J., Tortrat, F., Trepos, R., 2009. A decision-oriented model to evaluate the effect of land use an agricultural management on herbicide contamination in stream water. Environmental Modelling & Software 24, 1433–1446. doi:10.1016/j.envsoft.2009.06.002.

Gippel, C.J., 1995. Potential of turbidity monitoring for measuring the transport of suspended solids in streams. Hydrological Processes 9, 83–97. doi:10.1002/hyp.3360090108.

Glasgow, H.B., Burkholder, J.M., Reed, R.E., Lewitus, A.J., Kleinman, J.E., 2004. Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. Journal of Experimental Marine Biology and Ecology 300, 409–448. doi:10.1016/j.jembe.2004.02.022.

Grayson, R.B., Finlayson, B.L., Gippel, C.J., Hart, B.T., 1996. The potential of field turbidity measurements for the computation of total phosphorus and suspended solids loads. Journal of Environmental Management 47, 257–267. doi:10.1006/jema.1996.0051.

Grayson, R.B., Gippel, C.J., Finlayson, B.L., Hart, B.T., 1997. Catchment-wide impacts on water quality: the use of 'snapshot' sampling during stable flow. Journal of Hydrology 199, 121–134. doi:10.1016/S0022-1694(96)03275-1.

Harmel, R.D., Smith, D.R., King, K.W., Slade, R.M., 2009. Estimating storm discharge and water quality data uncertainty: a software tool for monitoring and modeling applications. Environmental Modelling & Software 24 (7), 832–842. doi:10.1016/j.envsoft.2008.12.006.

Hart, J.K., Martinez, K., 2006. Environmental sensor networks: a revolution in earth system science? Earth-Science Reviews 78, 177–191. doi:10.1016/j.earscirev.2006.05.001.

Helsel, D.R., 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley and Sons, New York, 250 pp.

Helsel, D.R., Hirsch, R.M., 2002. Statistical methods in water resources, Chapter A3. In: Techniques of Water-resources Investigations of the United States Geological Survey, Book 4, Hydrologic Analysis and Interpretation. United States Geological Survey, 523 pp. Available at: http://pubs.er.usgs.gov/usgspubs/twri/twri04A3.

Hill, D.J., Minsker, B., Amir, E., 2007. Real-time Bayesian anomaly detection for environmental sensor data. In: Proceedings of the 32nd Congress of the International Association of Hydraulic Engineering and Research (IAHR 2007), Venice, Italy. Available at: http://reason.cs.uiuc.edu/eyal/papers/Bayesian-anomaly-sensor-IAHR07.pdf.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resources Research 44, W05406. doi:10.1029/2007WR006392.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2009a. Components of an integrated environmental observatory information system. Computers & Geosciences (in review).

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009b. An integrated system for publishing environmental observations data. Environmental Modelling & Software 24, 879–888. doi:10.1016/j.envsoft.2009.01.002.

Johnes, P.J., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332, 241–258. doi:10.1016/j.jhydrol.2006.07.006.

Jordan, P., Arnscheidt, A., McGrogan, H., McCormick, S., 2007. Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser. Hydrology and Earth System Sciences 11 (1), 372–381.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42, W03S04. doi:10.1029/2005WR004362.

Kirchner, J.W., Feng, X., Neal, C., Robson, A.J., 2004. The fine structure of water-quality dynamics: the (high-frequency) wave of the future. Hydrological Processes 18, 1353–1359. doi:10.1002/hyp.5537.

Kronvang, B., Bruhn, A.J., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes 10, 1483–1501. doi:10.1002/(SICI)1099-1085(199611)10:11. <1483::AID-HYP386>3.0.CO;2-Y.

Liu, Y., Minsker, B., Hill, D., 2007. Cyberinfrastructure technologies to support QA/QC and event-driven analysis of distributed sensing data. In: Proceedings: International Workshop on Advances in Hydroinformatics, Niagara Falls, Canada Available at: http://colab.ncsa.uiuc.edu/EESHG1/Documents/repository/conference/CyberinfrastructureTechQAQC.pdf.

Maidment, D.R. (Ed.), 2008. CUAHSI Hydrologic Information System: Overview of Version 1.1, Consortium of Universities for the Advancement of Hydrologic Science Inc., Washington, DC, 92 pp. Available at: http://his.cuahsi.org/documents/HISOverview.pdf.

May, D.B., Sivakumar, M., 2009. Prediction of urban stormwater quality using artificial neural networks. Environmental Modelling & Software 24, 296–302. doi:10.1016/j.envsoft.2008.07.004.

McCuen, R.H., 2005. Hydrologic Analysis and Design, third ed.. Prentice Hall, Upper Saddle River, NJ, 888 pp.

Montgomery, J.L., Harmon, T., Kaiser, W., Sanderson, A., Haas, C.N., Hooper, R., Minsker, B., Schnoor, J., Clesceri, N.L., Graham, W., Brezonik, P., 2007. The WATERS Network: an integrated environmental observatory network for water research. Environmental Science and Technology 41 (19), 6642–6647. Available at: http://pubs.acs.org/subscribe/journals/esthag/41/i19/pdf/100107feature_waters.pdf.

Mourad, M., Bertrand-Krajewski, J.L., 2002. A method for automatic validation of long time series of data in urban hydrology. Water Science & Technology 45 (4–5), 263–270.

Nichols, M.H., Anson, E., 2008. Southwest watershed research center data access project. Water Resources Research 44, W05S03. doi:10.1029/2006WR005665.

Nolan, K.M., Bohman, L., Stamey, T., Firda, G., 2005. Stage–Discharge Relations – Basic Concepts Training Class USGS Scientific Investigations Report 2005–5028, USGS Training Class SW1409. Available at: http://wwwrcamnl.wr.usgs.gov/sws/SWTraining/RatingsWeb/Index.html.

Phillips, J.M., Webb, B.W., Walling, D.E., Leeks, G.J.L., 1999. Estimating the suspended sediment loads of rivers in the LOIS study area using infrequent samples. Hydrological Processes 13 (7), 1035–1050.

Quilbe, R., Rousseau, A.N., Duchemin, M., Poulin, A., Gangbazo, G., Villeneuve, J.P., 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: application to the Beaurivage River (Quebec, Canada). Journal of Hydrology 326, 295–310.

Rasmussen, T.J., Lee, C.J., Ziegler, A.C., 2008. Estimation of Constituent Concentrations, Loads, and Yields in Streams of Johnson County, Northeast Kansas, Using Continuous Water-quality Monitoring and Regression Models, October 2002 through December 2006, USGS Scientific Investigations Report 2008-5014, 103 pp. Available at: http://pubs.usgs.gov/sir/2008/5014/pdf/SIR2008-5014.pdf.

Rasmussen, T.J., Ziegler, A.C., Rasmussen, P.P., 2005. Estimation of Constituent Concentrations, Densities, Loads, and Yields in the Lower Kansas River, Northeastern Kansas, Using Regression Models and Continuous Water-quality Monitoring, January 2000 through December 2003, USGS Scientific Investigations Report 2005-5165, 117 pp. Available at: http://pubs.usgs.gov/sir/2005/5165/pdf/SIR20055165.pdf.

Robertson, D.M., Roerish, E.D., 1999. Influence of various water quality sampling strategies on load estimates for small streams. Water Resources Research 35 (12), 3747–3759.

Ryberg, K.R., 2006. Continuous Water-quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Red River of the North, Fargo, North Dakota, USGS Scientific Investigations Report 2006-5241, 35 pp. Available at: http://pubs.usgs.gov/sir/2006/5241/pdf/sir2006-5241.pdf.

Scholefield, D., Le Goff, T., Braven, J., Ebdon, L., Long, T., Butler, M., 2005. Concerted diurnal patterns in riverine nutrient concentrations and physical conditions. Science of the Total Environment 344, 201–210. doi:10.1016/j.scitotenv.2005.02.014.

Spackman Jones, A., 2008. Estimating Total Phosphorus and Total Suspended Solids Loads from High Frequency Data, M.S. thesis, Utah State University, Logan, UT, 144 pp. Available at: http://digitalcommons.usu.edu/etd/205.

Stubblefield, A.P., Reuter, J.E., Dahlgren, R.A., Goldman, C.R., 2007. Use of turbidometry to characterize suspended sediment and phosphorus fluxes in the Lake Tahoe basin, California, USA. Hydrological Processes 21 (3), 281–291. doi:10.1002/hyp.6234.

Tomlinson, M.S., De Carlo, E.H., 2003. The need for high resolution time series data to characterize Hawaiian streams. Journal of the American Water Resources Association (JAWRA) 39 (1), 113–123. doi:10.1111/j.1752-1688.2003.tb01565.x.

Uhrich, M.A., Bragg, H.M., 2003. 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. U.S. Department of the Interior, U.S. Geological Survey. Water-resources Investigations Report 03-4098. Available at: http://pubs.usgs.gov/wri/WRI03-4098/pdf/wri034098.pdf.

Vivoni, E.R., Camilli, R., 2003. Real-time streaming of environmental field data. Computers & Geosciences 29 (4), 457–468. doi:10.1016/S0098-3004(03)00022-0.

Wagner, R.J., Boulger Jr., R.W., Oblinger, C.J., Smith, B.A., 2006. Guidelines and Standard Procedures for Continuous Water-quality Monitors: Station Operation, Record Computation, and Data Reporting U.S. Geological Survey Techniques and Methods 1–D3, 51pp. + 8 attachments. Available at: http://pubs.water.usgs.gov/tm1d3.

WATERS Network, 2008. Draft Science, Education, and Design Strategy for the WATer and Environmental Research Systems Network. WATERS Network Project Office. Available at: http://www.watersnet.org/docs/SEDS-20080227-draft.pdf.

Wilkinson, S.N., Prosser, I.P., Rustomji, P., Read, A.M., 2009. Modelling and testing spatially distributed sediment budgets to relate erosion processes to sediment yields. Environmental Modelling & Software 24, 489–501. doi:10.1016/j.envsoft.2008.09.006.

Zaslavsky, I., Valentine, D., Whiteaker, T. (Eds.), 2007. CUAHSI WaterML, OGC Discussion Paper OGC 07-041r1, Version 0.3.0 Available at: http://www.opengeospatial.org/standards/dp.