SUSPENDED SEDIMENT AND TOTAL PHOSPHORUS LOADINGS
FROM SMALL AGRICULTURAL WATERSHEDS IN WESTERN ILLINOIS

BY

AMY M. RUSSELL

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Associate Professor Richard Cooke

ABSTRACT

On larger rivers, instantaneous sample concentrations are often treated as being representative of the mean daily concentration for load calculation purposes. This assumption, however, is not appropriate on streams where concentrations can change substantially within a day. In five small, rural watersheds in western Illinois, the collection and analysis of data during runoff events was done on a sub-daily time step. To have accurate load estimates, the selected load calculation method should correctly characterize loading behavior during short duration runoff events.

The use of statistical models with residuals-based error correction (i.e. the composite method) has become an increasingly popular technique for load calculations. This study is an application of error corrected regression models to compute continuous records of suspended sediment and total phosphorus concentrations at five watersheds in western Illinois. Due to the small drainage areas of the studied streams, all regression models were developed and applied using a 15-minute time-step. Four methods of constructing continuous concentration records were compared, and the best method to compute sediment and phosphorus loads for a 10-year period of study was identified.

For both suspended sediment and total phosphorus, load calculations by error corrected regression models produced estimates that were the most precise and least biased. Further, the method of error correction was not as critical as the act of error correction itself. During the ten-year study period, 5% of the record accounted for approximately 50% of the flow, 91% of the total phosphorus load and more than 96% of the sediment load. On average, the 1-day maximum accounted for 10% of annual flow and more than 30% of annual sediment and phosphorus loads.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Dr. Richard Cooke for his encouragement, patience, and thoughtful critiques of this research work. I would like to thank my committee members Dr. Prasanta Kalita and Dr. George Czapar for their opinions and advice. I would also like to thank Vern Knapp and Dr. Momcilo Markus of the Illinois State Water Survey for their questions, suggestions, and comments as I refined my thesis topic.

This research would not have been possible without the amazing datasets collected by the Illinois State Water Survey. I would like to thank Jim Slowikowski for not simply training his field crews but for also being vigilant in his oversight and instilling in them the importance of collecting data with the same level of integrity the 10,000th time as is required the 1st time. As part of their primary duties for the Illinois State Water Survey, Kip Stevenson, Ted Snider, Josh Stevens, and Mike Smith were the dedicated staff responsible for streamgaging and collecting the water quality samples used in this study. I sincerely thank them all.

I would also like to thank my family, friends, and colleagues at the Illinois State Water Survey who endured this long process with me. Finally, I would like to thank my partner, Jim, for shouldering far more than his fair share while I pursued this degree. This would not have been possible without him.

TABLE OF CONTENTS

1. INTRODUCTION

Excessive sediment and phosphorus contribute to degradation of aquatic habitat. In fact, suspended sediment/solids and total phosphorus are the leading impairments of Illinois inland lakes according to the Illinois Environmental Protection Agency (2012, 2010). Sedimentation and total phosphorus are also two of the top ten impairments of Illinois streams.

Monitoring streamflow and water quality is critical to developing effective strategies for reducing sediment and phosphorus loads. Suspended sediment and phosphorus transport in rural Illinois watersheds is primarily runoff driven. Since most of the sediment and phosphorus mass is delivered during runoff events, intensive monitoring programs typically involve fixed-period sampling (weekly, monthly) supplemented with storm event sampling. Detailed records of concentration and flow are needed to compute loadings for a stream.

To investigate the role of individual runoff events on loads, a load calculation method must be selected that allows for investigating small time scales. Statistical models with residuals-based error correction (i.e. the composite method) have become an increasingly popular technique for load calculations. This study is an application of error corrected regression models to compute continuous records of suspended sediment concentration (SSC) and total phosphorus (TP) concentration. Due to the small drainage area of the five studied streams, all regression models were developed and applied using a 15-minute time-step. Four methods of constructing continuous concentration records were compared (linear interpolation of samples, regression model, regression model with traditional composite method, regression model with modified composite method). The best method was selected in order to compute sediment and phosphorus loads for a 10-year period of study.

2. Objectives

The main hypothesis of this research is that on small, rural streams the overwhelming majority of loadings of sediment and phosphorus occur during storm flow events; due to the short duration of the storms, the collection and analysis of data during these runoff events must be done on a sub-daily time step. Further, the technique used to calculate loads at these sites should consider flow, season, and hydrographic position. If storm sampling is a component of the monitoring plan, the use of observed data to adjust load calculations greatly improves load estimates. The specific objectives were to:

1. Develop multiple linear regression models for study sites and identify best model for estimating SSC and TP concentrations. Test the hypothesis that a regression model used for TP and SSC load calculations in small watersheds with fixed interval and storm event samples will require multiple explanatory variables.
2. Explore error correction techniques for improving concentration estimates and subsequent load calculations
3. Determine proportion of sediment and phosphorus loading as function of time and flow.

3. BACKGROUND

Load estimation is often a critical element in developing solutions to address water quality impairments. In order to effectively address these concerns, some information is needed regarding the location and delivery of the sources of non-point source (NPS) pollution. This is typically accomplished by monitoring various locations within a watershed to determine loads, the mass of pollutant passing a location during a time period of interest. Estimating a stream's load is achieved by measuring the streamflow, collecting water quality samples to determine constituent concentrations, and selecting a method for computing loads.

3.1 FIELD METHODS FOR MONITORING SEDIMENT AND PHOSPHORUS LOADS IN SMALL, RURAL WATERSHEDS

3.1.1 Streamgaging

The number of small, rural watersheds gaged in Illinois has decreased substantially over the past 40 years. According to Knapp and Markus (2003), in 1971 the U.S. Geological Survey (USGS) operated 20 rural gages in Illinois with a drainage area of less than 30 square miles and by 2001, only one such gage remained. Within the Illinois River basin, the number of USGS gages in rural watersheds less than 500 square miles decreased from 26 to 9 over the same time period.

Because of this lack of streamflow information, NPS monitoring efforts of small, rural watersheds in Illinois typically are initiated with little knowledge of the flow characteristics or typical solute responses of the streams prior to the start of data collection. Several factors, such as the rapid changes in stage, high stream velocities and substantial debris flows during storm events, contribute to the difficulties in monitoring small streams, specifically in western Illinois.

These, as well as other factors inherent in small watersheds, need to be taken into consideration when selecting stage sensing equipment (Slowikowski et al., 2003).

Stage records are typically produced by continuous monitoring of stream levels using stage sensing equipment, such as a pressure transducer, bubbler or radar unit. Converting these stage records to continuous records of streamflow in rural streams is obtained through rigorous streamgaging. A sufficient number of discharge measurements collected throughout the entire range of stages in all seasons is required to develop relationships between stage and discharge (rating curves). Additional analyses are required to determine shifts, or small, usually temporary changes to the rating due to conditions such as vegetative growth in the channel, leaf build-up in the streams during low flows, and scour or deposition following substantial flow events. Further estimates of streamflow are also required based on conditions of ice cover or periods of backwater due to beaver dams, log jams, etc. All of these above-mentioned impacts are more pronounced on small streams.

Stage readings are recorded on a sub-daily time-step, typically every 15 minutes, and then the stage-discharge rating curve is applied to this record to produce a "continuous" record of streamflow. The mean daily discharge is then computed by averaging these instantaneous streamflow values each day. The USGS is the primary agency responsible for streamgaging in the nation, and the standard deliverables of the USGS streamgaging program are records of mean daily streamflow, although there has been a recent push to also release the instantaneous record.

This lack of existing streamflow information greatly impacts the water quality sampling design of a monitoring program as well. Obtaining samples representative of flow conditions is extremely challenging prior to the establishment of a rating curve for the site. Without gage records, an understanding of the stream’s flashiness cannot be quantified and is difficult to incorporate into sampling protocols. These measures of stream response are typically some of the very questions attempting to be answered by the monitoring studies themselves.

3.1.2 Water Quality Sampling

There are several methods for collecting stream water quality samples. Most samples can be considered discrete, composite or continuous. A discrete sample is a single, instantaneous sample collected at a single location in the stream cross-section. Composite samples are combinations of several discrete samples and can be flow-weighted, time-weighted, or spatially integrated. Flow-weighted and time-weighted samples are typically collected with the assistance of an automated pump sampler that can be programmed to sample based on different conditions. Spatially integrated samples are samples manually collected, typically with a depth-integrated sampler, at one or more verticals within the stream. Continuous measures of water quality are collected using in-situ sensors and are most often used to collect parameters such as temperature, pH, dissolved oxygen, and specific conductance. Turbidity sensors are also available and have been the focus of recent research to investigate the use of turbidity as a surrogate for simulating continuous records of sediment concentrations (Bragg and Uhrich, 2010; Williamson and Crawford, 2011).

The most robust monitoring programs include a combination of manual samples, cross-section composite samples, depth-integrated samples, and automated pump samples. Depth-integrated samplers are most often used for sediment and other particulate constituents. Because these constituents are not evenly distributed within a stream column, depth-integrated samplers are used to collect a representative flow-weighted sample (Edwards and Glysson, 1999). Cross-section samples are collected on larger streams with significant mixing zones, although even on small streams it is good practice to do routine cross-section composites to verify the assertion that a point sample from a single vertical within the stream is representative of concentrations across the entire stream cross-section.

3.1.3 SSC versus TSS

Often the terms suspended sediment concentration (SSC) and total suspended solids (TSS) are used interchangeably, but these two measures of suspended sediment/solids actually refer to different analytical methods. Both methods determine the concentration of sediment in water, but suspended sediment concentration (SSC) is determined by filtering and drying an entire water sample, while TSS analysis uses an aliquot. Because the act of sub-sampling the bulk water sample during TSS analysis can result in an underrepresentation of larger particle sizes, SSC has been deemed a more accurate measure, especially when high sand content is present (Gray et al., 2000). It is a common misconception that TSS includes organic matter while SSC does not. This is not the case as both methods involve only filtering and oven drying. To determine the organic portion of a sample, an additional subsequent analysis would be required where the oven dried material is then placed in a furnace to allow the organic portion to burn off. The portion lost is considered the volatile suspended fraction.

3.2 METHODS FOR COMPUTING SEDIMENT AND PHOSPHORUS LOADS

Once a monitoring effort has a streamflow record and water quality samples, the next step is to decide on the load calculation method. To truly compute a stream's load would require a continuous record of streamflow (hydrograph) and continuous record of concentration data (chemograph). Multiplying these two records together, with the appropriate conversion factor, produces a continuous record of flux or loading rate. The summation of these loading rates over the time period of interest produces the total load.

Because concentration data are typically not available at the same resolution as streamflow records (i.e. “continuous” or 15-minute records), different algorithms have been developed for estimating loads. An important consideration when selecting a suitable load algorithm is how each method determines representative concentrations during un-sampled or under-sampled periods. Furthermore, despite the availability of 15-minute records of streamflow, many traditional load calculation methods utilize mean daily discharge, as that is the main product of USGS gaging.

Birgand et al. (2010) investigated the role of load algorithm selection and various sampling frequencies in the uncertainty of annual nitrate loads. They found that the choice of algorithm significantly influenced the accuracy and precision of nitrate load estimates. Their research, however, largely focused on load methods commonly used with sparse data sets, so they only evaluated various averaging and interpolation methods, which are most often used with infrequent sampling programs. They did not evaluate any regression methods for load calculation. This study, on the other hand, will investigate load estimation techniques commonly used with intensive monitoring programs and large data sets.

Common types of algorithms for computing sediment loads for intensively monitored streams typically fall into the following four categories.

- Sediment rating curve
- Worked record
- Regression method
- Composite method

The method selected often depends on availability of data and the purpose of the analysis.

3.2.1 Sediment Rating Curve

Sediment rating curves (or sediment transport curves) are developed by plotting sediment concentration (mg/L) or sediment flux (tons/day) as a function of streamflow at the time of sampling. The most common best-fit line is typically the power function

$$C = aQ^b \tag{3.1}$$

or log-transforming the variables results in an equation of the form

$$\ln C = a + b \ln Q \tag{3.2}$$

where C is concentration, Q is discharge and a, b are regression coefficients. The above equations can also be written with sediment flux as the dependent variable.

In situations where the relationship between flow and sediment discharge is not best described by a linear relationship, adjustments can be made to better fit the non-linear relationship (Crowder et al., 2007; Demissie et al., 2003). Simon et al. (2004) describe the technique of breaking the sediment rating curve into multiple equations based on flow strata and using professional judgment to visually determine the equation break points. To account for the hysteresis effect seen on some streams, the USGS also supports developing separate sediment discharge ratings for a site based on whether samples were collected on the rising or falling limb of the hydrograph (Glysson, 1987). The development of separate sediment discharge rating curves is also encouraged by the USGS when a site exhibits strong seasonal variation in sediment concentrations.

Sediment rating curves are typically used to estimate sediment concentration or sediment flux at established streamgaging locations where sediment samples have been collected on a limited basis over several years, and detailed sediment records are not available. Horowitz

(2003) found that for long-term data sets the development of a single sediment rating curve for the entire period of record can be quite accurate, although the development of annual sediment rating curves can produce somewhat better load estimates. Sediment rating curves, adjusted for non-linearities, were used in the sediment budget developed for the Illinois River (Demissie et al., 2003), which found that the Spoon and Sangamon watersheds are the highest sediment producing watersheds in the Illinois River basin.

3.2.2 Worked Record

The worked record approach refers to USGS recommended practices for developing a continuous time-series record of concentration based on observed concentrations supplemented with estimated concentrations inserted to improve the shape of the chemograph constructed by piece-wise linear interpolation of observations (Porterfield, 1972). This approach is typically used at established streamgaging locations where more extensive sediment sampling has been conducted and is also sometimes referred to as the mass accumulation or integration method in the literature (Haggard et al, 2003; Robertson, 2003; Robertson and Roerisch, 1999).

Porterfield cautions that the estimation of concentration data should be conducted by "personnel with knowledge of the sampling program, the physical and cultural environments affecting the stream regimen and sediment sources, and the fundamentals of sediment transport." While the use professional judgment is emphasized, this method is extremely subjective, especially during extended periods of missing concentration data. To facilitate the development of these temporal concentration graphs, the USGS developed and endorses the use of the Graphical Constituent Loading and Analysis System (GCLAS). GCLAS provides a graphical suite of tools to visualize and explore various relationships such as sediment transport curves plotted beside the corresponding hydrographs and chemographs (McKallip et al, 2001).

This approach is typically used with sediment data but could be applied to any constituent that exhibits a strong enough relationship to discharge, season, or hydrographic position that missing data can be estimated with some confidence. Because of the high level of knowledge of the stream system required to employ this method as well as the reliance on extensive observed data, this method is typically considered the most accurate load estimation method, and loads computed by this approach are often used as “true” measures of load in studies exploring optimizing sampling frequency or comparing other load calculation methods (Robertson, 2003; Robertson and Roerisch, 1999).

3.2.3 Regression Method

The load estimation method of choice for most large-scale nutrient studies is multiple linear regression because of its ability to incorporate terms for a variety of explanatory variables beyond simply discharge which are often required when investigating the cycling and transport of nutrients. Additionally, in large-scale assessments of stream loads it is unrealistic that one researcher can have the intimate knowledge of each study site that is required to implement the worked record approach for load estimation. Furthermore, rarely are sufficient nutrient samples available to construct the continuous chemograph.

The most common multiple regression model used in nutrient studies is Cohn’s 7-parameter regression model (Cohn et al., 1992) which takes the form:

$$\ln(C) = \beta_0 + \beta_1 \ln Q + \beta_2 \ln Q^2 + \beta_3 \sin(2\pi T) + \beta_4 (\cos 2\pi T) + \beta_5 T + \beta_6 T^2 \qquad \textbf{(3.3)}$$

where C is the constituent concentration, Q is the discharge rate, T is decimal time (time measured in years) and β are model coefficients. This model is one of many equations available within the USGS-developed LOAD ESTimator (LOADEST) program for estimating constituent loads (Runkel et al., 2004).

Guo et al. (2002) utilized this 7-parameter regression model to evaluate the effects of sampling frequency and monitoring duration on nitrate load estimates for a site in central Illinois. This regression model was also used in the 1999 Gulf of Mexico hypoxia study (Goolsby et al.), as well as subsequent studies of nutrient loads within the Mississippi-Atchafalaya River basin (Aulenbach et al., 2007). Terrio (2006) used the following 4-parameter regression model (a modified version of Equation 3.3) to determine constituent fluxes in the Illinois River basin.

$$\ln(CQ) = \beta_0 + \beta_1 \ln Q + \beta_2 \sin 2\pi T + \beta_3 \cos 2\pi T \qquad \textbf{(3.4)}$$

The variables to explain long-term trends in the data were most likely omitted because of the relatively short period of study (four years).

Wang and Linker (2008) proposed the addition of two additional terms to Equation 3.3 to simulate the clockwise hysteresis often seen where sediment concentrations on the rising limb of a storm hydrograph are higher than concentrations measured at the same discharge on the falling limb (Robertson, 2003; Richards et al., 2001).

In all of the above cited literature, the discharge term used in the regression models was the mean daily discharge. For nutrient load estimation an underlying assumption is often that an instantaneous sample concentration is representative of the mean daily concentration at a given site. This is often referred to as the "big rivers" modeling approach and is not suitable on flashy streams where concentrations can rise and fall several orders of magnitude during a runoff event.

3.2.4 Composite Method

The composite method describes a load estimation technique that uses observed concentrations to improve the concentrations predicted by regression methods. In addition to predicting a continuous concentration record, regression models also generate a mixed-frequency dataset of residual concentrations (the difference between observed and simulated concentrations). The composite method, as described by Aulenbach and Hooper (2006), creates a continuous record of residuals through linear interpolation. These residuals are then used to adjust the regression model predicted concentrations, which in effect sets the model predicted concentrations equal to the observed concentrations. This approach is most appropriate when the observed data exhibits strong serial correlation.

The composite method can be used with regression models of any format. Aulenbach and Hooper (2006) used a hyperbolic function to relate solute concentration to discharge as this described a two-component mixing model which was deemed most appropriate for their dissolved constituents of interest (alkalinity and chloride). Vanni et al. (2001) presented a method for estimating hourly nitrogen and phosphorus concentrations between observed samples. Their Q-proportionate method, which used residual concentrations computed from constituent rating curves (Equation 3.2) to determine the amount of error correction between observations, is equivalent to the composite method.

As part of a recent study, Verma et al. (2012) developed four alternative error correction techniques and tested their performance using daily records of nitrate-N concentration for two large watersheds in central Illinois. The new error correction techniques modified both the temporal distribution of residual concentrations, as well as the measure of the residual itself. To construct the residual concentration curves, the authors inserted vertices at mid-points between each observation. For a triangular distribution, the residuals were set equal to zero at these mid-points, and the remaining residuals were determined by linear interpolation between the observations and these mid-points. For a rectangular distribution, the residual concentration at each observation was maintained for the interval extending from the previous mid-point to the subsequent mid-point. In addition to computing the residual as the difference between observed and modeled concentrations, the authors also proposed computing the residual as the proportion of the observed concentration to the modeled concentration. The four error correction techniques (triangular residual, rectangular residual, triangular proportional, and rectangular proportional) were compared to the traditional composite method of error correction (linear interpolation of residuals), and the rectangular proportional (RP) method was found to perform the best for their study sites.

3.3 Study Site description

The data used in this study were collected by the Illinois State Water Survey between 1999 and 2009 at five sites within the Spoon and Sangamon watersheds (Demissie et al., 2001). The locations of the watersheds within the Illinois River basin are shown in Figure 3-1.

Drainage areas of the five study sites are provided in Table 3-1 along with selected watershed characteristics computed for the gaged portions of these watersheds. The two smallest study watersheds are located adjacent to each other in the lower Sangamon River basin (Figure 3-2). The North Creek watershed is located wholly within the Court Creek watershed, which is adjacent to the Haw Creek watershed (Figure 3-3). While some land use characteristics differ among the sites, general soils and physiography of the watersheds are similar.

It should be noted that the Site IDs used in this study were assigned in order of increasing drainage area.

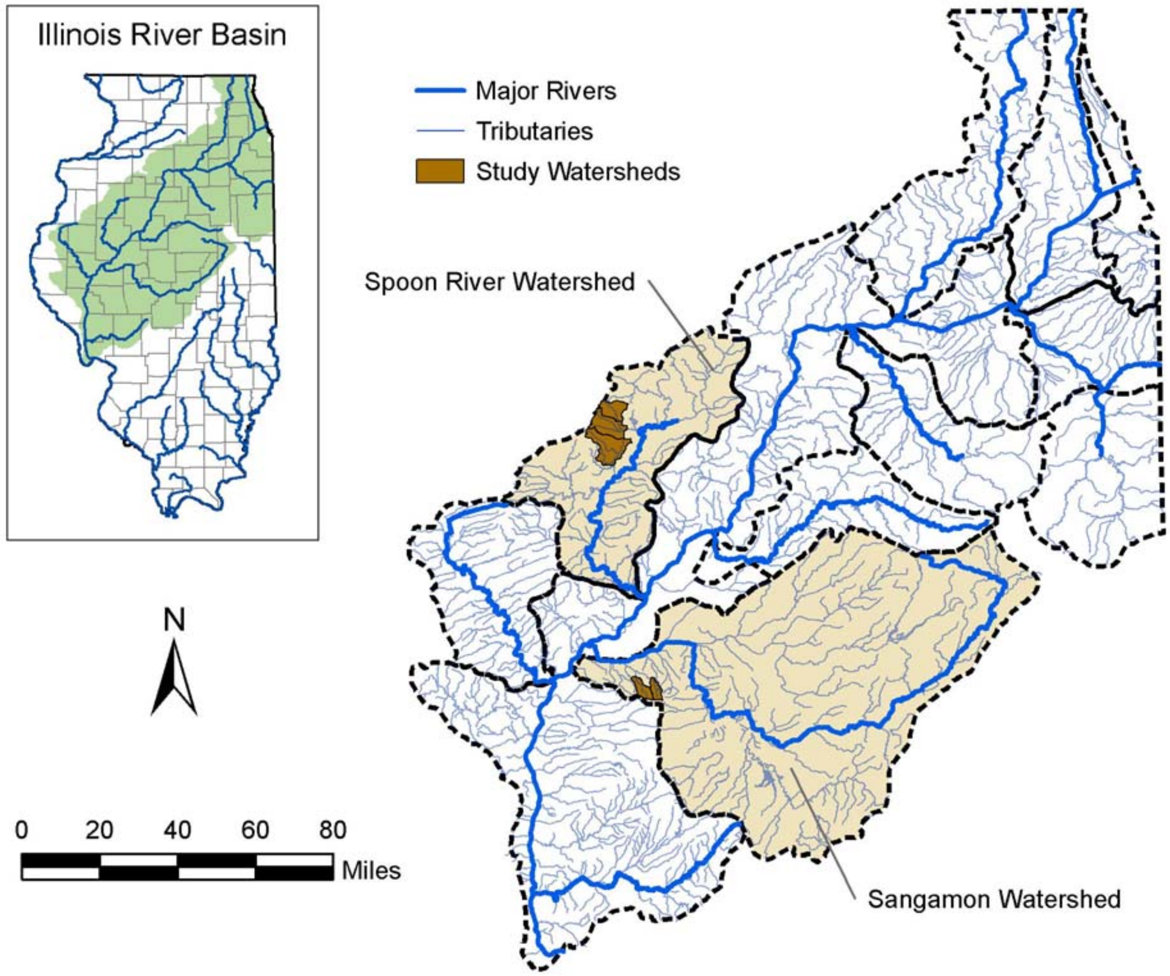

Figure 3-1. Location of study watersheds within the Illinois River basin

Table 3-1. Study sites and watershed characteristics

Site ID	Site Name	Drainage Area (sq mi)	Stream Slope (ft/mi)	Land Use (Percent of Study Watershed)			Major Watershed
				Agriculture	Forest	Other	
1	Cox Creek	11.7	13.6	93	6	1	Sangamon
2	Panther Creek	16.5	13.3	75	22	3	Sangamon
3	North Creek	26.6	21.1	65	31	4	Spoon
4	Haw Creek	55.3	8.3	80	14	6	Spoon
5	Court Creek	67.4	14.5	69	23	8	Spoon

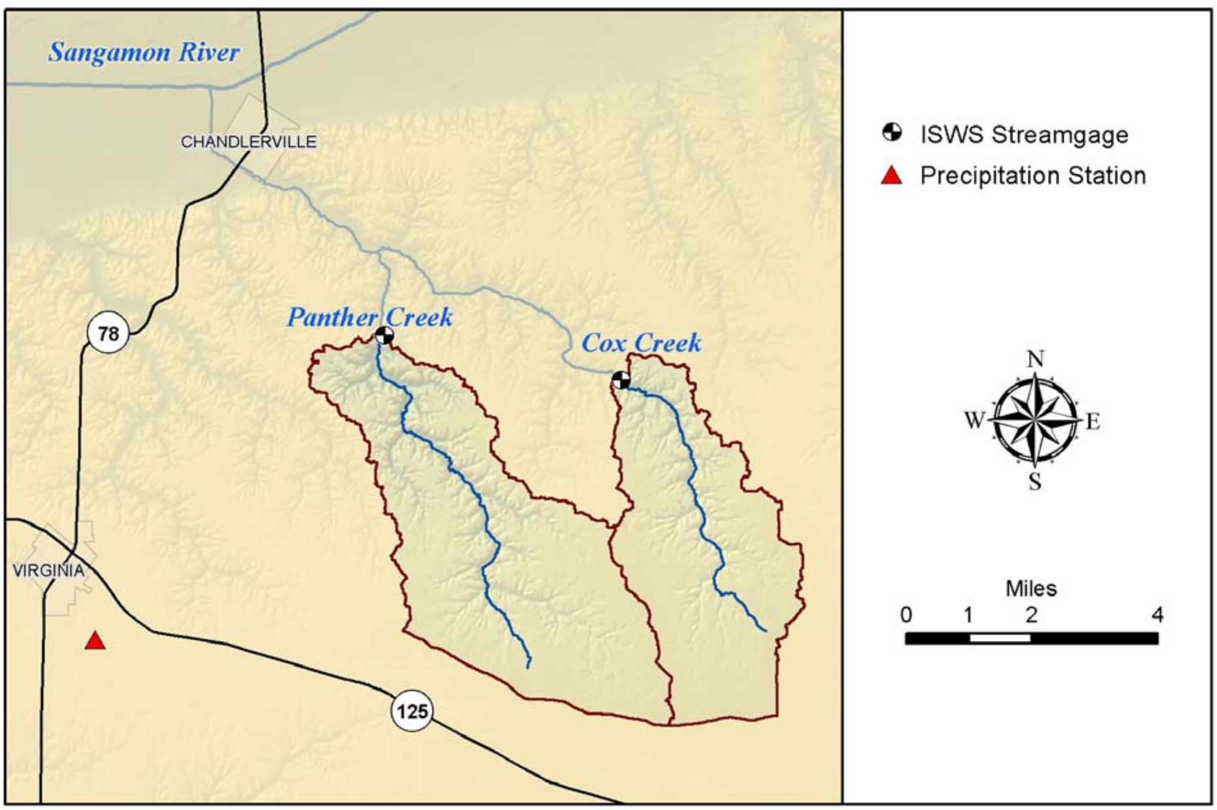

Figure 3-2. Location of study sites within the Sangamon River watershed

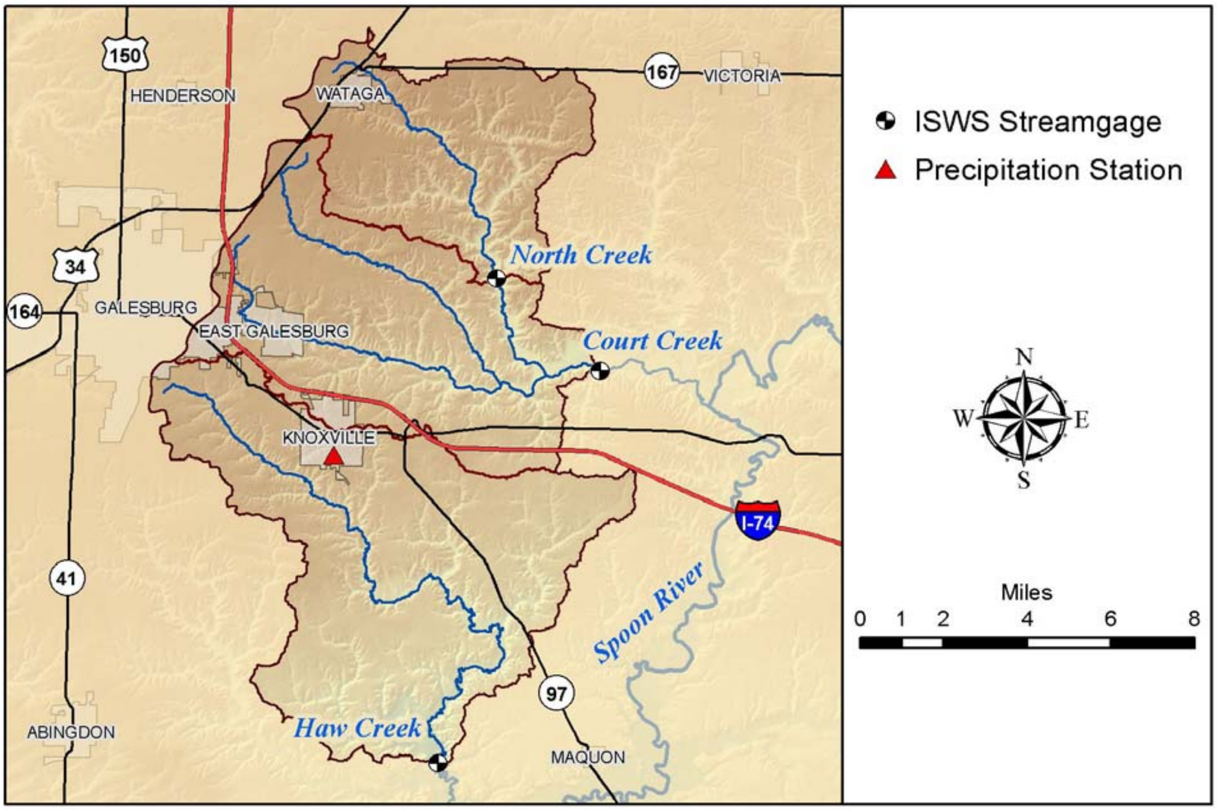

Figure 3-3. Location of study sites within the Spoon River watershed

Due to their relative proximity to each other, the variability in precipitation patterns at the study sites should be minimal.

3.4 Period of Study

The data used in this study were collected during the period October 1, 1999–September 30, 2009 (Water Years 2000 – 2009).

3.4.1 Precipitation

The nearest National Weather Service COOP precipitation stations are located in Virginia, IL and Knoxville, IL. The Virginia gage is located southwest of Panther Creek and Cox Creek (Figure 3-2), while the Knoxville gage is located within the Haw Creek watershed

(Figure 3-3). Precipitation data during the 10-year period of study are summarized in Table 3-2 for these two nearby rain gages. To place this period of study in the proper historical context for this region of Illinois, it is helpful to compare the precipitation during the period of study to the long-term (30-year) average for these rain gages. At the Knoxville gage, Water Years 2002 and 2007 essentially experienced normal precipitation, while six out of ten years reported below average precipitation, and only the last two years of the study period (WY 2008-2009) were above normal. While the annual precipitation totals at the Virginia gage were often several inches more or less than Knoxville's totals, the two gages followed similar patterns of above or below normal precipitation as evident in Figure 3-4. While Water Years 2008 and 2009 were wet years at Virginia as well, the precipitation surplus in 2008 was more than 13", a much more pronounced departure than the 5" surplus recorded at Knoxville.

Table 3-2. Annual precipitation totals during study period, inches

	Knoxville	Virginia
WY2000	35.89	33.28
WY2001	37.76	37.47
WY2002	40.66	43.45
WY2003	31.95	34.13
WY2004	39.14	35.57
WY2005	35.48	35.06
WY2006	34.43	31.02
WY2007	40.49	26.98*
WY2008	45.60	53.11
WY2009	49.74	47.54
10-yr Average	39.11	37.76*
1981-2010 Normal	40.62	39.75

*Note: Includes 81 days of Missing Data

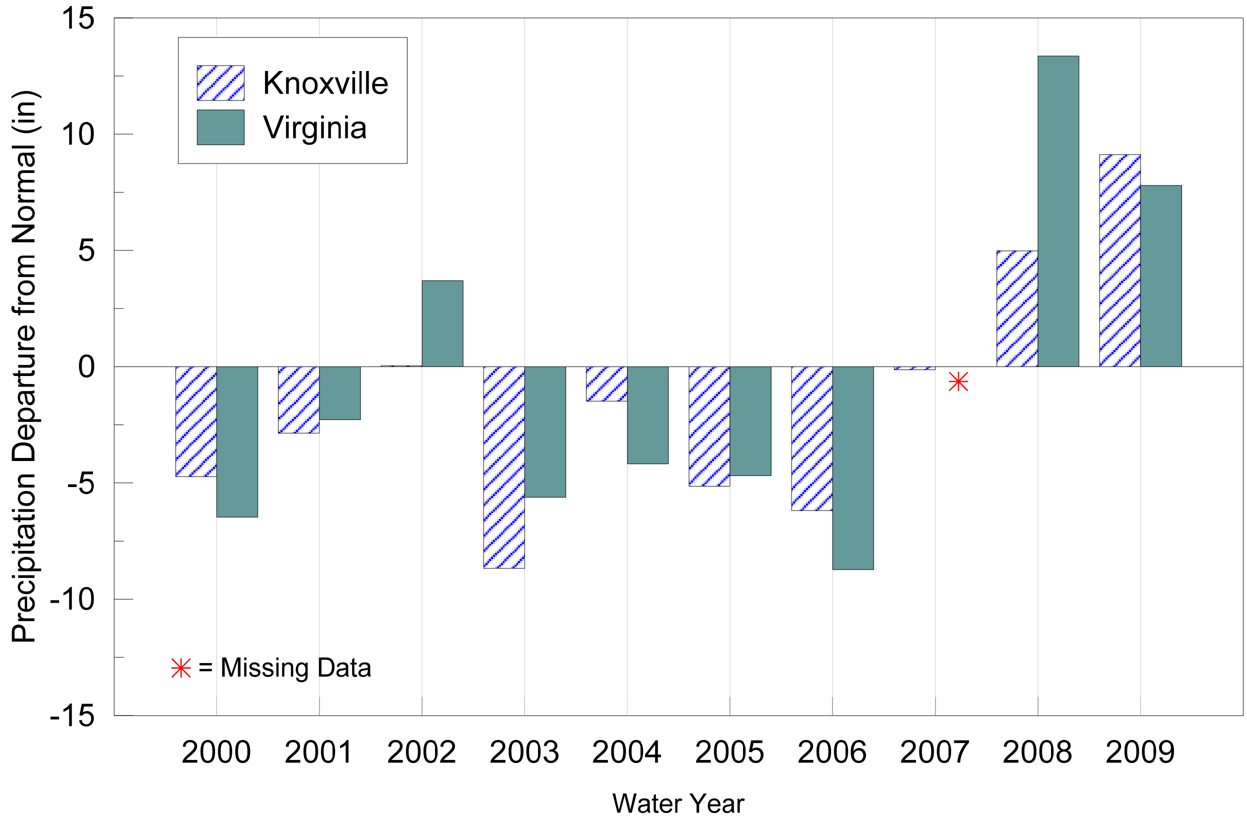

Figure 3-4. Annual precipitation during study period, as compared to 1981-2010 average

Water Year 2005 was generally considered a drought year, however, October 2004 – January 2005 had a greater than 6" surplus of precipitation at both gages during those four months, but the remainder of the year, as well as 2006, was quite dry. Water Year 2008 was the wettest year of the study period at the Virginia gage. Each month from June-Sept 2008 were 1-5" above normal. The Virginia gage reported more than 8" of precipitation in September 2008, while Knoxville reported nearly 10" that month. The most rainfall occurred at the Knoxville gage during Water Year 2009 with more than a 9" surplus for the year, and February–June 2009 were all 1-4" above normal for precipitation.

3.4.2 Streamflow

When investigating sediment and nutrient loadings, it is important to have a firm understanding of the streamflow conditions during the study period. The presence of unusually wet or dry years can have significant implications on the interpretation of annual loads. Determining whether annual streamflow was largely the result of a single, large storm event or several smaller storm events can be equally important.

Annual Streamflow Variation

Annual flows for the five ISWS streamgages are summarized in Table 3-3. Normalizing the streamflow as inches of runoff illustrates the similar annual flows for these streams by removing the effect of differences due to drainage area. The variation in annual runoff during the 10-year study period is presented in Figure 3-5. As the precipitation records (Figure 3-4) would suggest, 2008 and 2009 experienced the highest flows of the study period, and total streamflow was least in 2006 at all study sites, except Haw Creek (2003). In general the flows are similar at the sites, but differences in 2008 can be attributed to the much higher precipitation totals that year in the Sangamon watersheds, as compared to the Spoon watersheds.

Table 3-3. Annual flow statistics for ISWS gages, Water Year 2000-2009

Site ID	Site Name	Mean Annual Flow (cfs)	Mean Annual Runoff (in)
1	Cox Creek	9.0	10.4
2	Panther Creek	12	9.6
3	North Creek	18	9.3
4	Haw Creek	39	9.6
5	Court Creek	46	9.3

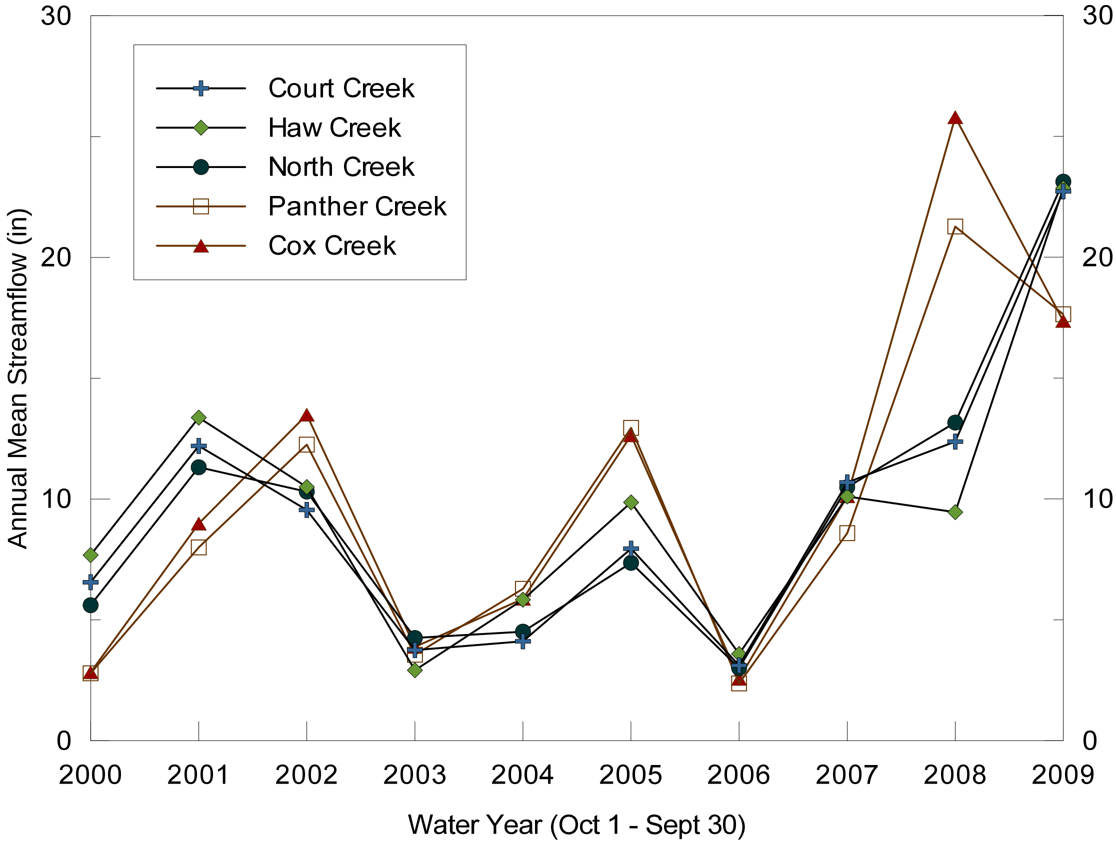

Figure 3-5. Annual runoff at study sites, Water Year 2000-2009

Seasonal Streamflow Variation

In order to explore the seasonal variability of streamflow at the study sites, flow values for each month were averaged to determine the mean monthly streamflow for each month and year. These monthly flows during the 10-year study period were then ranked, and the maximum, median and minimum monthly streamflow values for Panther Creek and Court Creek are provided in Figure 3-6 and Figure 3-7, respectively.

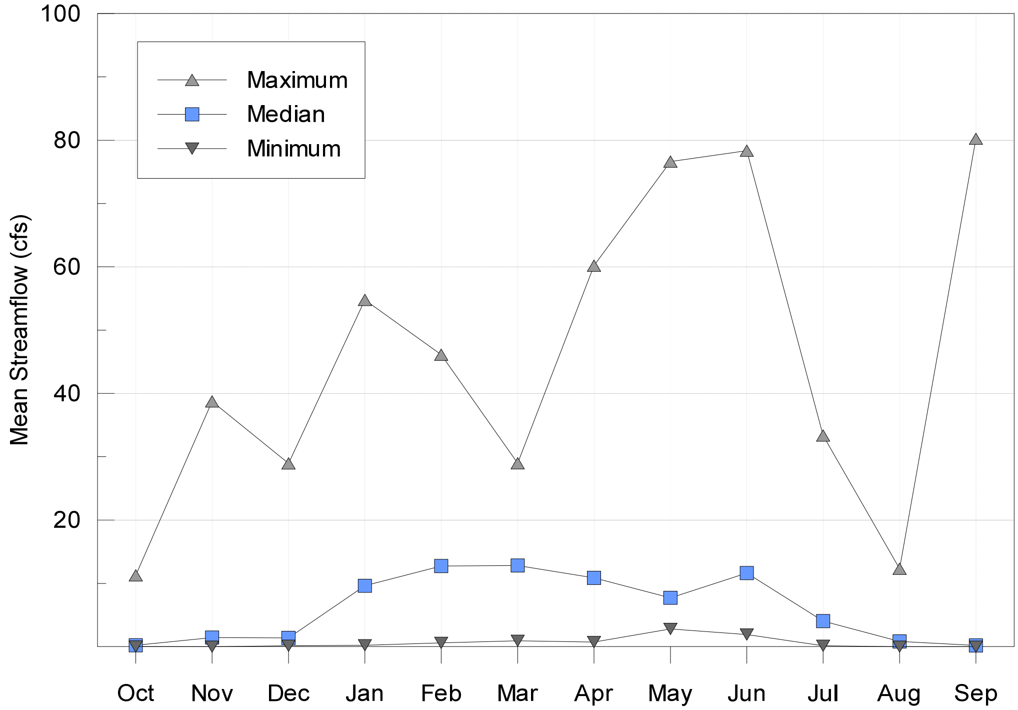

Figure 3-6. Monthly Streamflow at Panther Creek (Site #2), Water Year 2000-2009

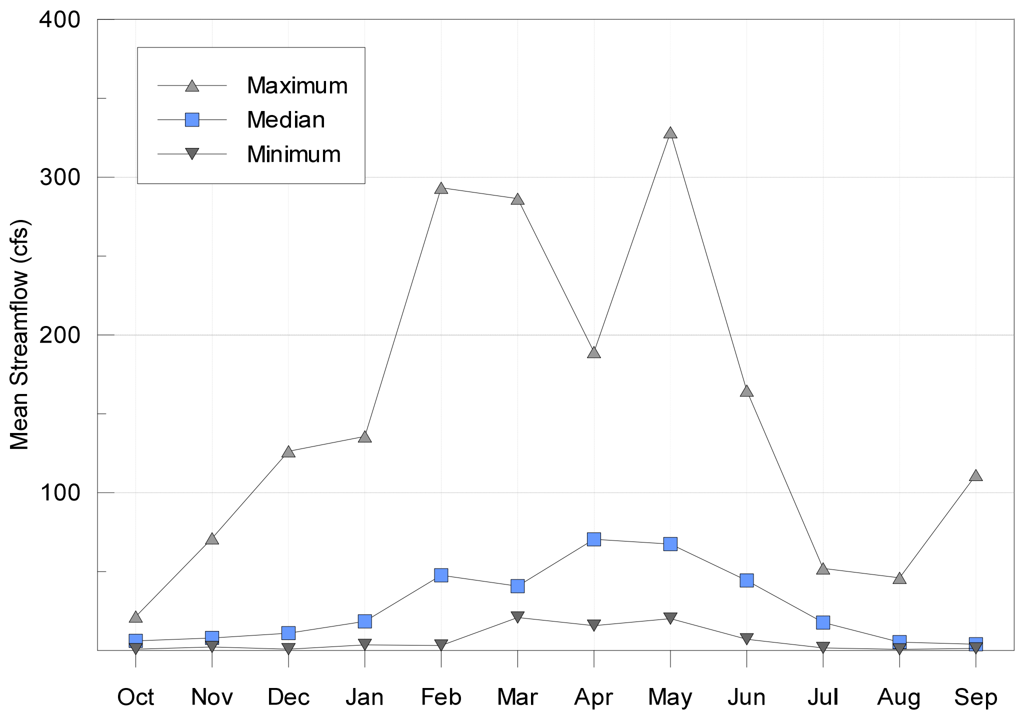

Figure 3-7. Monthly streamflow at Court Creek (Site #5), Water Year 2000-2009

The ratios of monthly to annual flows during the 10-year study period are presented in Table 3-4. Flows are typically greatest in the months of April-June in the Sangamon watersheds (Sites 1 and 2) and Feb-May in the Spoon watersheds (Sites 3-5), while flows are the lowest during Aug-Nov at all study sites. The magnitude of the variation in monthly flows is greater at the Spoon watershed study sites than the Sangamon sites, as evidenced by the fact that Feb-May flows in the Spoon watershed are nearly twice their annual flows, while Aug-Nov flows are approximately one-fifth of their annual flows.

Table 3-4. Ratio of monthly flow to annual flow, Water Year 2000-2009

Site ID	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1	0.19	0.52	0.70	1.39	1.44	1.07	1.54	1.61	1.61	0.76	0.23	0.93
2	0.20	0.53	0.71	1.25	1.55	1.13	1.65	1.53	1.61	0.66	0.26	0.93
3	0.13	0.24	0.66	0.68	1.61	2.00	1.84	2.29	1.49	0.51	0.21	0.36
4	0.27	0.40	0.74	0.78	1.88	1.62	1.66	2.19	1.59	0.43	0.20	0.24
5	0.17	0.32	0.68	0.82	1.82	1.86	1.63	2.11	1.52	0.49	0.25	0.33

During the study period, the months with the highest total flow were September 2008 at sites 1 and 2 in the lower Sangamon, followed closely by May 2002 and June 2008. In the Spoon watershed, the highest monthly flows were recorded in March or May 2009, and February 2001 experienced the next highest flows outside of 2009 at these three sites. The lowest recorded flows during the study period occurred in 2005 and 2007. In fact, Sites 1-3 all experienced extended periods of zero flow during this study.

Flow Duration Analysis

Flow duration curves display the percent of time flows were equaled or exceeded during a given time period. In order to illustrate the high instantaneous flows experienced at these sites, that would not be apparent from the mean daily flow record, flow duration curves for the study sites were developed by sorting and ranking each 15-minute reading of streamflow recorded during the 10-year study period. These values were then plotted on probability paper (Figure 3-8). The range of flows experienced at these five sites varies over 5-6 orders of magnitude and show similar shapes/characteristics. Cox, Panther, and North Creek experienced periods of no flow 5-7% of the time during the study period.

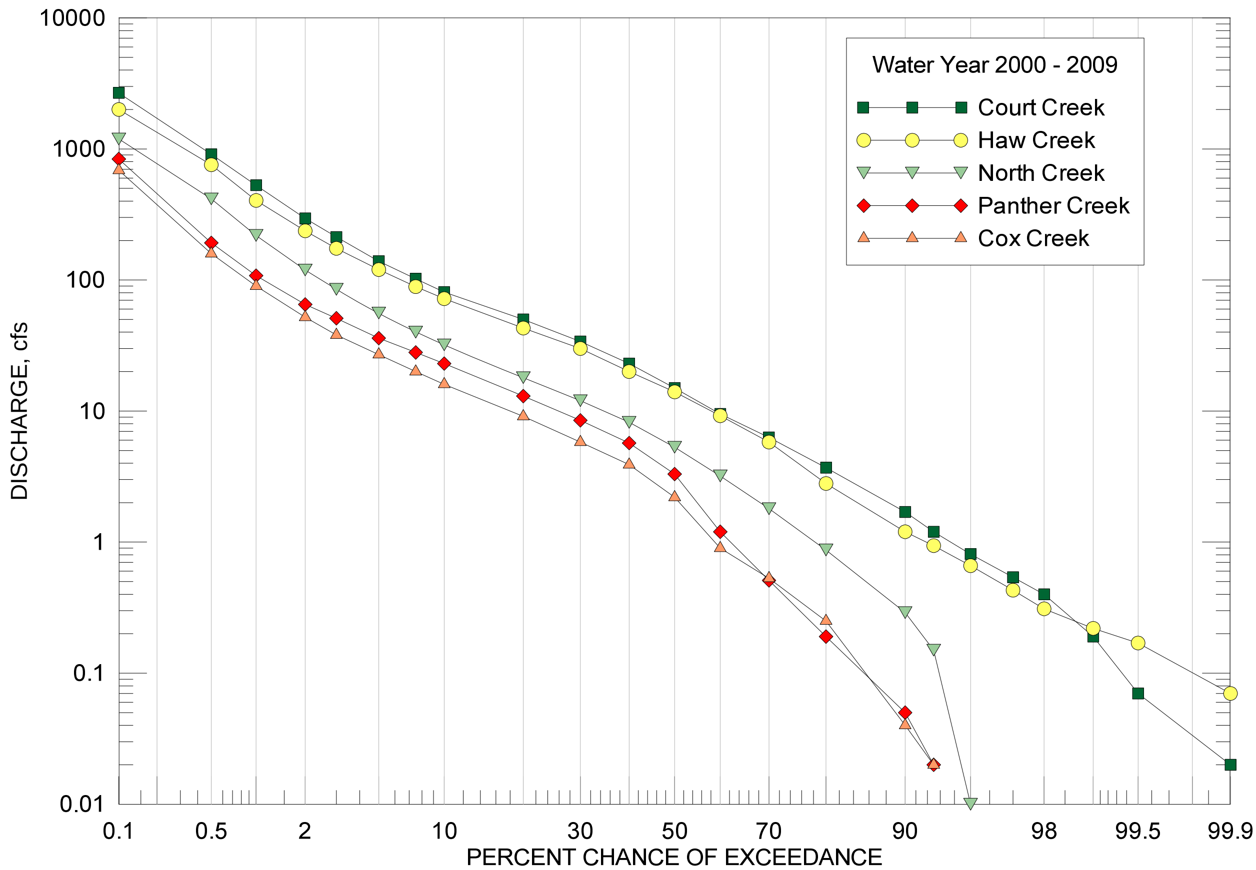

Figure 3-8. Flow duration curves for study sites

Haw Creek exhibited slightly higher extreme low flows than Court Creek, the larger adjacent watershed. This can be attributed to the fact that the Knoxville STP (NPDES ID: IL0022209) discharges into an unnamed tributary to Haw Creek approximately 13 miles upstream of the Haw Creek gage. This facility's average monthly discharge during the study period varied from 0.3-3.0 cfs, amounts small enough to be imperceptible in the flow duration curves during most flow conditions. However, during periods of extreme low flow, it appears that Haw Creek would probably be a dry stream if not for the discharge of the Knoxville STP.

Flashiness

Stream flashiness refers to the rate of change in streamflow and the frequency of these changes. Differences in stream flashiness can be attributed to soils, geology, land use, drainage area, and presence of point sources. An index developed by Baker et al. (2004) quantifies the flashiness of a stream by summing the absolute values of changes in streamflow and then dividing by the total of all mean daily discharges during the period of interest. This index is most commonly computed using mean daily streamflow records, but can be used with records of any regular time-step. The Richards-Baker flashiness index is defined as the sum of the absolute value of changes in discharge divided by the sum of the flows for the period of interest, as shown in Equation 3.5.

$$\textbf{R-B index} = \frac{\Delta t \sum_{i=1}^{n} | q_i - q_{i-1} |}{\sum_{i=1}^{n} q_i} \qquad \textbf{(3.5)}$$

where q is flow at any regular time-step t (measured in days).

The Richards-Baker Index (RBI) of flashiness was computed for ISWS gaging stations using mean daily, hourly, and 15-minute records of streamflow (Table 3-5). The streamflow response at Cox Creek was the flashiest of all study sites. Many of the runoff events at this site were measured in hours. Typically the larger watersheds have much slower responses to storm events. As a point of comparison, Verma et al. (2012) reported RBI values (based on mean daily flows) of approximately 0.2 for the Sangamon River at Monticello, and Baker et al. (2004) reported an RBI of 0.266 for the Spoon River at London Mills computed for Water Year 1975-2001.

Table 3-5. RBI of flashiness of streams at study sites, Water Year 2000-2009

Site ID	Site Name	Major Watershed	15-min Record	Hourly Record	Mean Daily Flows
1	Cox Creek	Sangamon	3.391	3.169	0.713
2	Panther Creek	Sangamon	2.647	2.466	0.638
3	North Creek	Spoon	2.053	1.944	0.667
4	Haw Creek	Spoon	1.163	1.110	0.554
5	Court Creek	Spoon	1.769	1.649	0.604

Stream flashiness is an important flow characteristic to evaluate, because many water quality constituents experience rapid changes in concentration during these periods of rapid changes in streamflow, specifically during the rising limb of an event. For many small rural streams, the loadings of particulate constituents during these large flow events of short duration can comprise a majority of the annual load. In *A Study of Measurement and Analysis of Sediment Loads in Streams*, the Federal Interagency Sedimentation Project (FISP, 1940) investigated the suspended sediment loading characteristics of small streams and found that for eleven small streams in the Midwest the 1-day maximum load experienced during a15-month monitoring period represented 8-36% of the total load. While conservation tillage has increased and fertilizer usage has decreased since this early study, more recent studies still support the finding that a few high flow events can account for the overwhelming majority of non-point source loadings of particulate constituents such as SSC, TSS, and TP (Markus and Demissie, 2006; Royer et al., 2006; Haggard et al., 2003, Richards et al., 2001).

On streams the size of our study sites, using mean daily flow to characterize flashiness or stream response does not capture the rapidity of changes in the stream hydrographs. However, the ratios of flashiness indices computed using the different resolutions of streamflow data (Table 3-6) indicate that using hourly instead of 15-minute streamflow record would capture most of the oscillations in flow seen at the finer time-scale. To illustrate the potential information lost by using mean daily records, a graph comparing Panther Creek's 15-minute streamflow record to its mean daily flow record for a period of 30 days in 2002 is provided in Figure 3-9. Since mean daily flow is computed from the higher resolution data, the total flow volume is equivalent with either record, but the rates of change in streamflow will not be adequately captured using the mean daily flow record for these small streams.

Table 3-6. Effect of using more detailed streamflow record to compute flashiness of streams at study sites

Site ID	Site Name	Drainage Area (sq mi)	Ratio of 15-min/Hourly RBI	Ratio of 15-min/Daily RBI
1	Cox Creek	11.7	1.07	4.76
2	Panther Creek	16.5	1.07	4.15
3	North Creek	26.6	1.06	3.08
4	Haw Creek	55.3	1.05	2.10
5	Court Creek	67.4	1.07	2.93

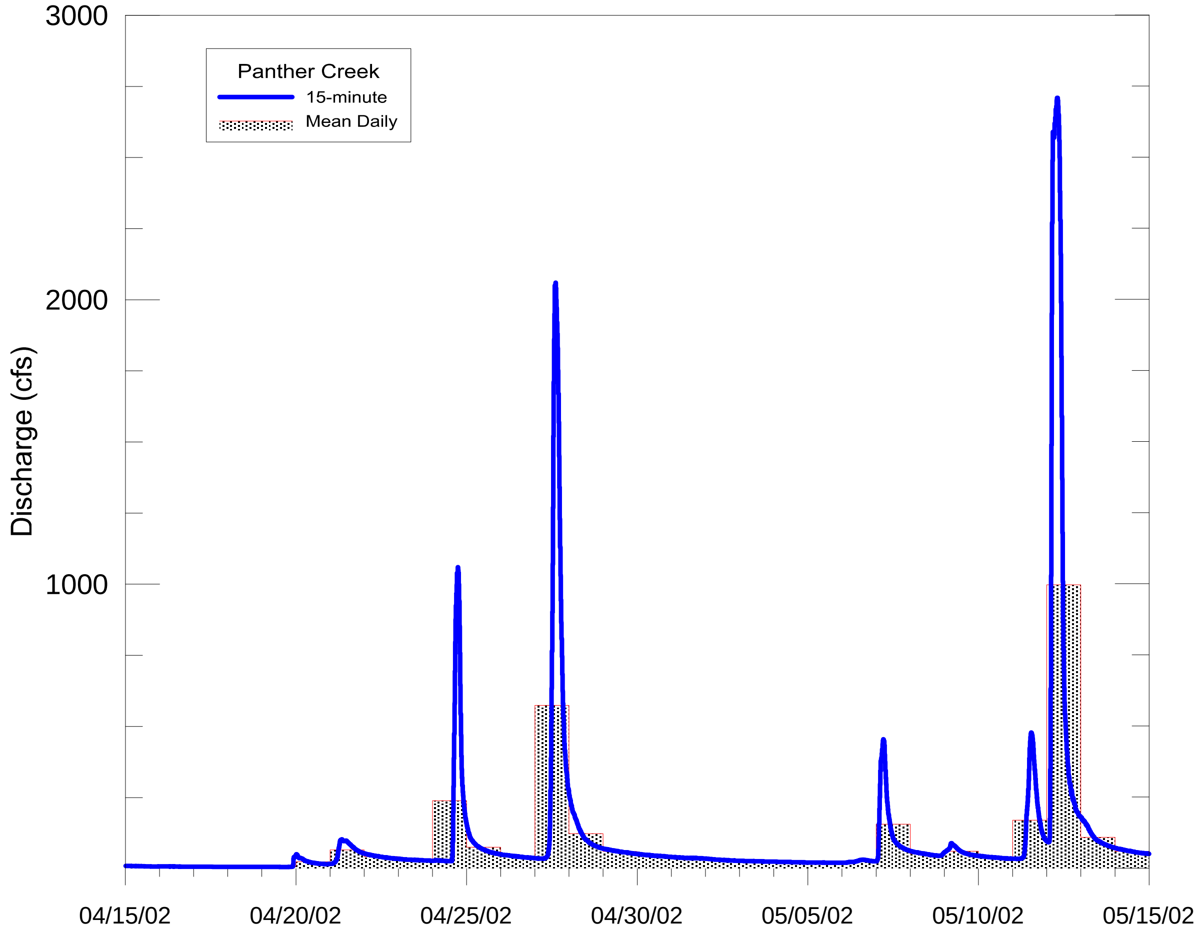

Figure 3-9. Comparison of 15-minute streamflow record to mean daily for Panther Creek, 4/15/2002-5/15/2002

3.4.3 Sediment and Phosphorus Concentrations

Samples used in this study were collected through a combination of routine and event sampling. All field methods followed USGS protocols and are detailed in the project QAPP (Demissie et al., 2000). Depth-integrated sediment samples were collected on a weekly basis and during storm events. A peristaltic pump sampler was used to collect additional samples during storm events based on characteristics of each site's stage hydrograph (Slowikowski et al., 2003). Pump samplers were also used to collect daily sediment samples during all but the lowest flow conditions.

Depth-integrated phosphorus samples were collected on a monthly basis and during storm events. Samples obtained from the pump sampler were occasionally submitted for TP analysis but only if the samples could be preserved with sulfuric acid and put on ice within 24 hours of sample collection.

The average number of samples collected annually at each site during the period of study is summarized in Table 3-7. The fewer TP samples at the Sangamon watershed sites (Panther and Cox Creek) reflect the impact of stream flashiness and the resulting difficulty in obtaining samples during storm events of short duration. Sediment samples are collected much more frequently than TP samples due to the lower expense of analyzing them and the less restrictive holding time. As a result of the monthly frequency for the routine sampling of total phosphorus as well as the infrequent use of pump samplers for TP sample collection, there were approximately ten times more sediment samples than TP samples collected during the study period.

Table 3-7. Average number of samples per year by site

Site Number	Site Name	Sediment	Phosphorus
1	Cox Creek	296	29
2	Panther Creek	372	29
3	North Creek	414	38
4	Haw Creek	405	39
5	Court Creek	331	39
Period of Record		12/1999 – 9/2009	3/2000 – 9/2009

Box and whisker plots summarizing the range of sediment and phosphorus concentrations observed at the study sites is provided in Figure 3-10 and Figure 3-11, respectively. Like streamflow, sediment and phosphorus concentrations vary over several orders of magnitude. Observed sediment concentrations ranged from a high of more than 48,000 mg/L at Panther Creek (Site #2) to less than 1 mg/L at North Creek (Site #3). Observed TP concentrations ranged from 11.2 mg/L at Panther Creek to levels below the method detection limit at all sites. Over the course of the 10-year study period, the MDL for TP analyses varied from 0.03 to 0.09 mg/L. For those samples with concentrations not detected at levels above the MDL, the MDL in place at the time was used as the observed concentration. Due to the low percentage of non-detects in the total set of observed samples, this was an appropriate approach for handling censored data.

The sediment concentration quartiles shown in Figure 3-10 are also provided in tabular form (Table 3-8), along with additional percentiles of interest. This concentration percentile information should be interpreted with care due to the heavy influence of storm samples on these data sets. These summary statistics describe the distribution of samples collected but do not necessarily reflect the true sample population and are most likely biased high. For example, in addition to the targeting of storm events, daily sediment samples were not obtained during extreme low flow conditions.

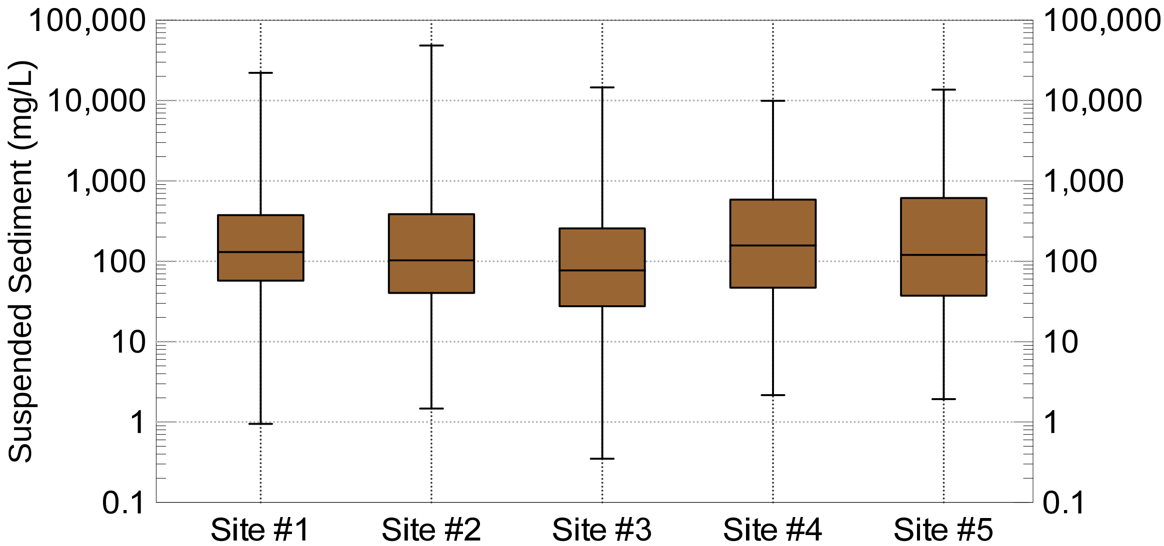

Figure 3-10. Observed sediment concentrations at study sites, December 1999-September 2009

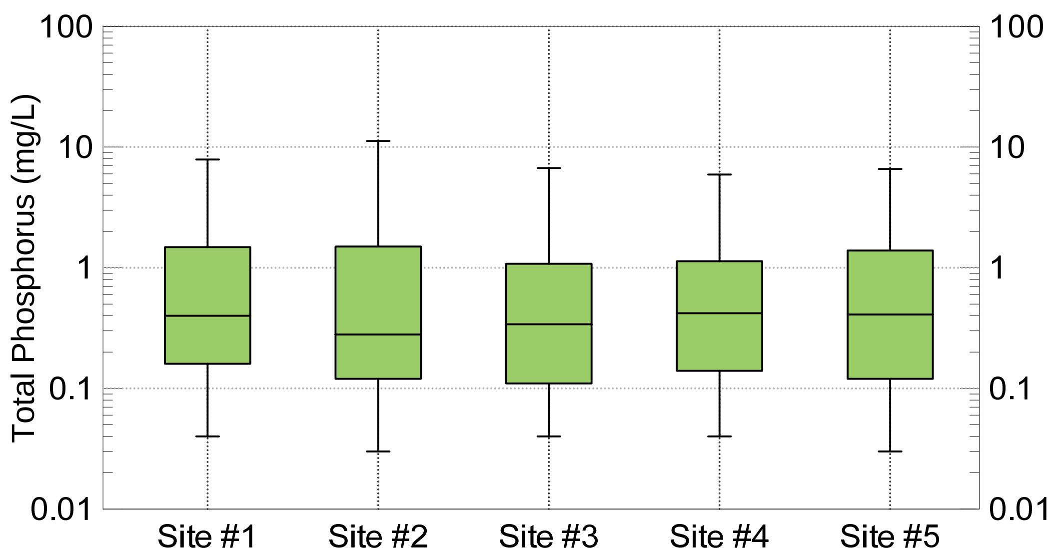

Figure 3-11. Observed phosphorus concentrations at study sites, March 2000-September 2009

Table 3-8. Statistical summaries of suspended sediment concentration (mg/L) at study sties, WY2000-2009

	Site #1	Site #2	Site #3	Site #4	Site #5
Maximum	22,067	48,289	14,565	9,879	13,632
99th Percentile	9,336	11,012	6,298	5,185	6,928
90th Percentile	2,134	2,244	1,234	1,616	2,176
75th Percentile	375	385	256	584	611
Median	130	102	77	157	120
25th Percentile	57	40	28	47	37
10th Percentile	27	18	12	15	13
1st Percentile	8	5	4	4	4
Minimum	1	1	0.4	2	2
Count	2910	3658	4069	3984	3258

The total phosphorus concentration percentiles are presented in Table 3-9. Once again this concentration percentile information is heavily influenced by storm samples. In fact only 115 samples at each site were collected as routine monthly samples, meaning 60-70% of all observed TP samples were collected during runoff events.

Table 3-9. Statistical summaries of total phosphorus concentration (mg/L) at study sties, WY2000-2009

	Site #1	Site #2	Site #3	Site #4	Site #5
Maximum	7.90	11.21	6.69	5.92	6.58
99th Percentile	7.02	7.05	5.75	4.50	5.93
90th Percentile	3.61	3.30	2.46	2.31	2.76
75th Percentile	1.48	1.50	1.06	1.13	1.38
Median	0.40	0.26	0.34	0.42	0.41
25th Percentile	0.16	0.12	0.11	0.14	0.12
10th Percentile	0.08	0.08	0.06	0.08	0.07
1st Percentile	0.05	0.05	0.04	0.05	0.04
Minimum	0.04	0.03	0.04	0.04	0.03
Count	275	276	365	371	369
% Non-Detects	3%	5%	7%	3%	7%

The monthly variation in sediment concentrations is shown in Figure 3-12 and Figure 3-13 for the Sangamon sites and Spoon sites, respectively. Samples collected in October-December at Cox Creek and Panther Creek exhibit lower concentrations than those collected in other months but this may be due in large part to the fact that fewer samples were collected October-December due to the lower flows these months. Also the pump sampler cannot be used in extreme cold weather or when a substantial portion of the stream is frozen. Higher concentrations in January and February are a result of the higher percentage of samples collected during storm events. The number of samples collected each month for each study site is summarized in Table 3-10. At the Spoon watershed sites, some of the highest sediment concentrations were observed in February.

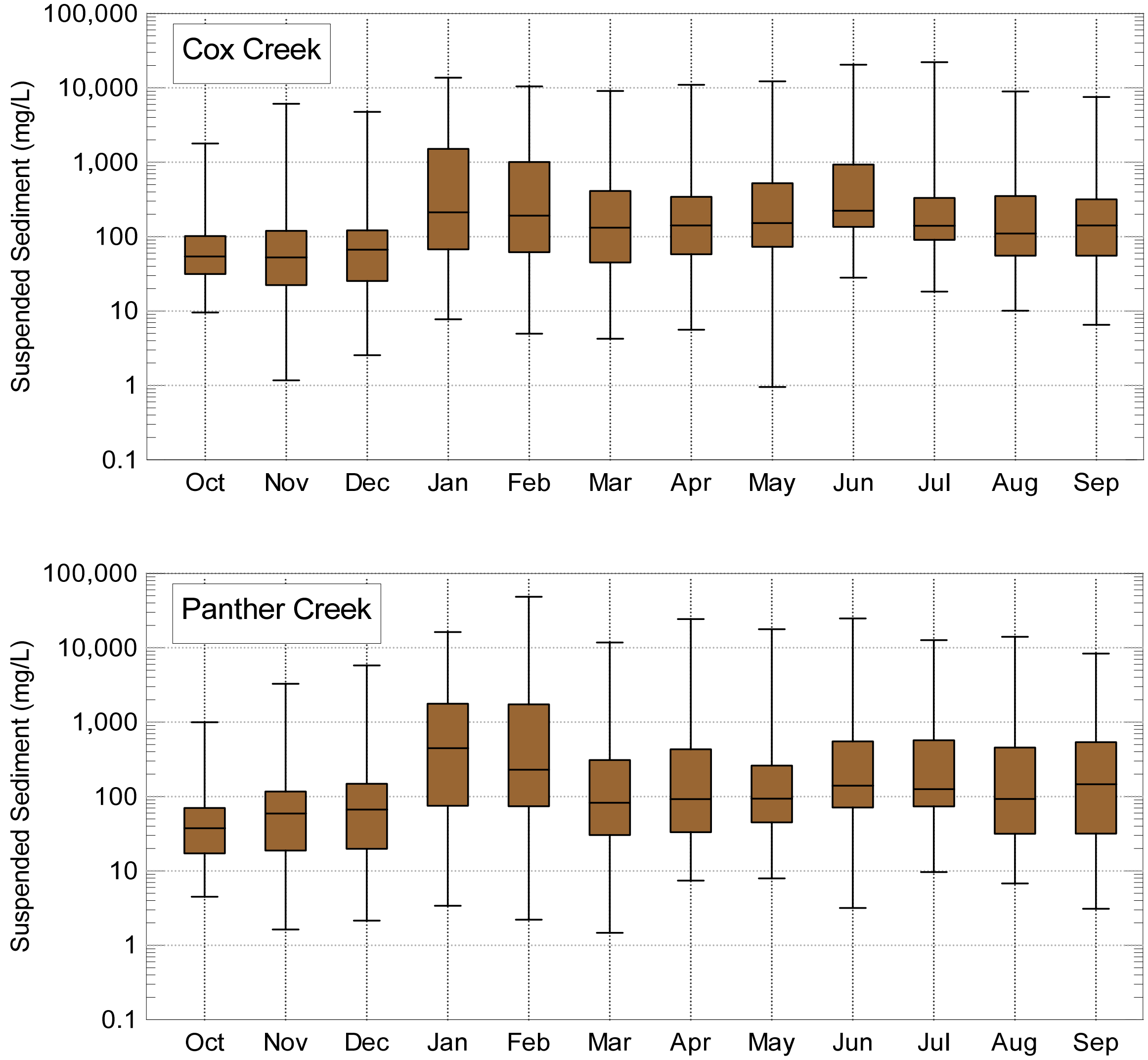

Figure 3-12. Monthly variation in sediment concentrations at Sangamon sites

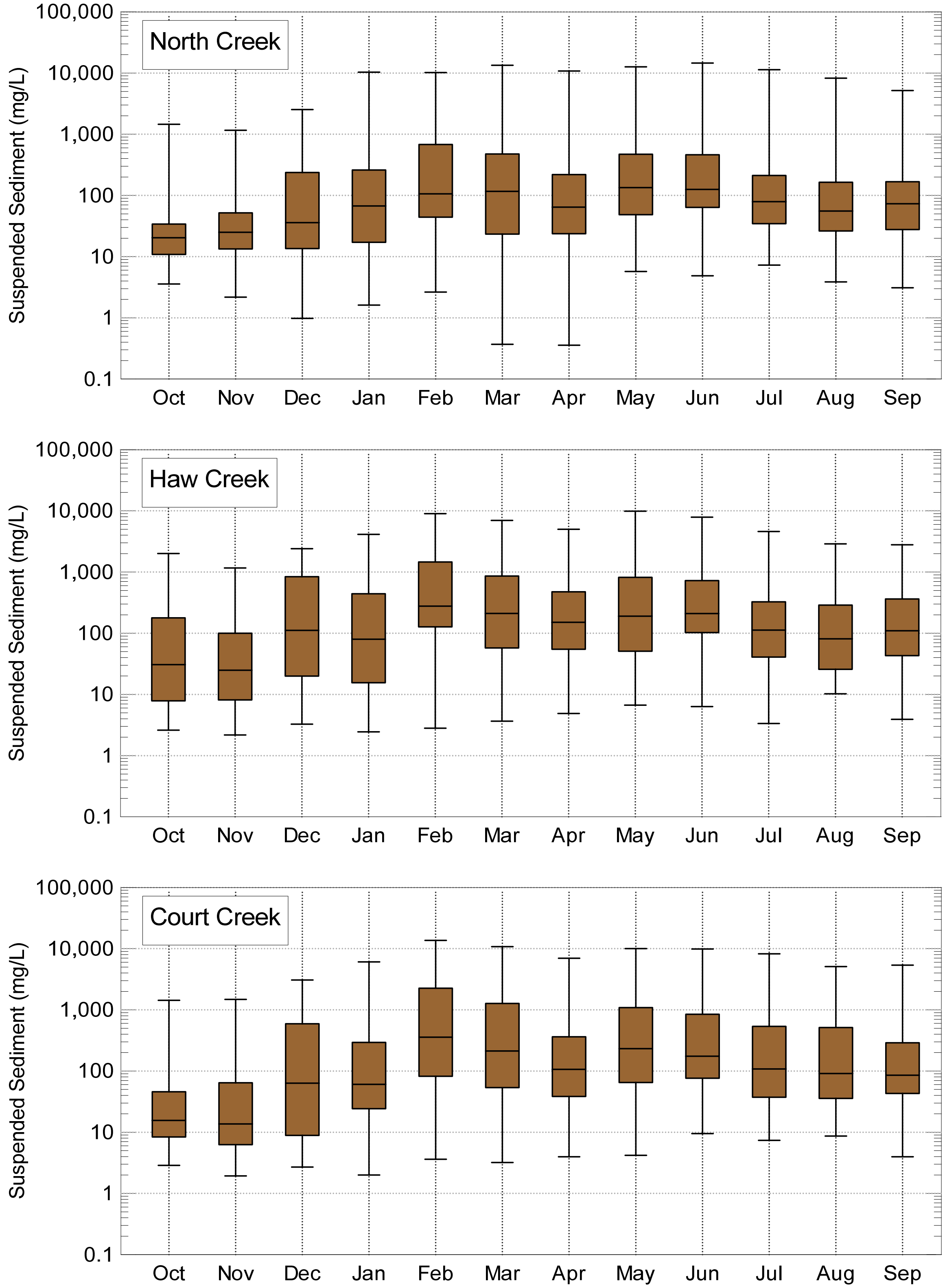

Figure 3-13. Monthly variation in sediment concentrations at Spoon sites

Table 3-10. Monthly distribution of sediment samples collected by site

Site Number	10	11	12	1	2	3	4	5	6	7	8	9	Total Number of Samples
1	144	231	154	155	194	310	350	317	321	281	219	234	2910
2	183	288	199	185	255	354	447	484	480	296	230	257	3658
3	240	207	158	140	199	423	550	668	539	389	269	287	4069
4	171	161	150	169	245	462	620	691	703	284	143	185	3984
5	103	174	105	156	173	374	464	568	477	310	200	154	3258

The monthly variation in total phosphorus concentrations is shown in Figure 3-14 and Figure 3-15 for the Sangamon and Spoon sites, respectively. When looking at these figures, it is important to remember the number of samples collected each month is not equal (Table 3-11). The effect of storm samples on concentration distributions is even more pronounced with the TP data. During the entire 10-year period of study, in the month of October no TP storm samples were collected at North or Court Creek and only 1-3 total storm samples were collected that month at the other study sites.

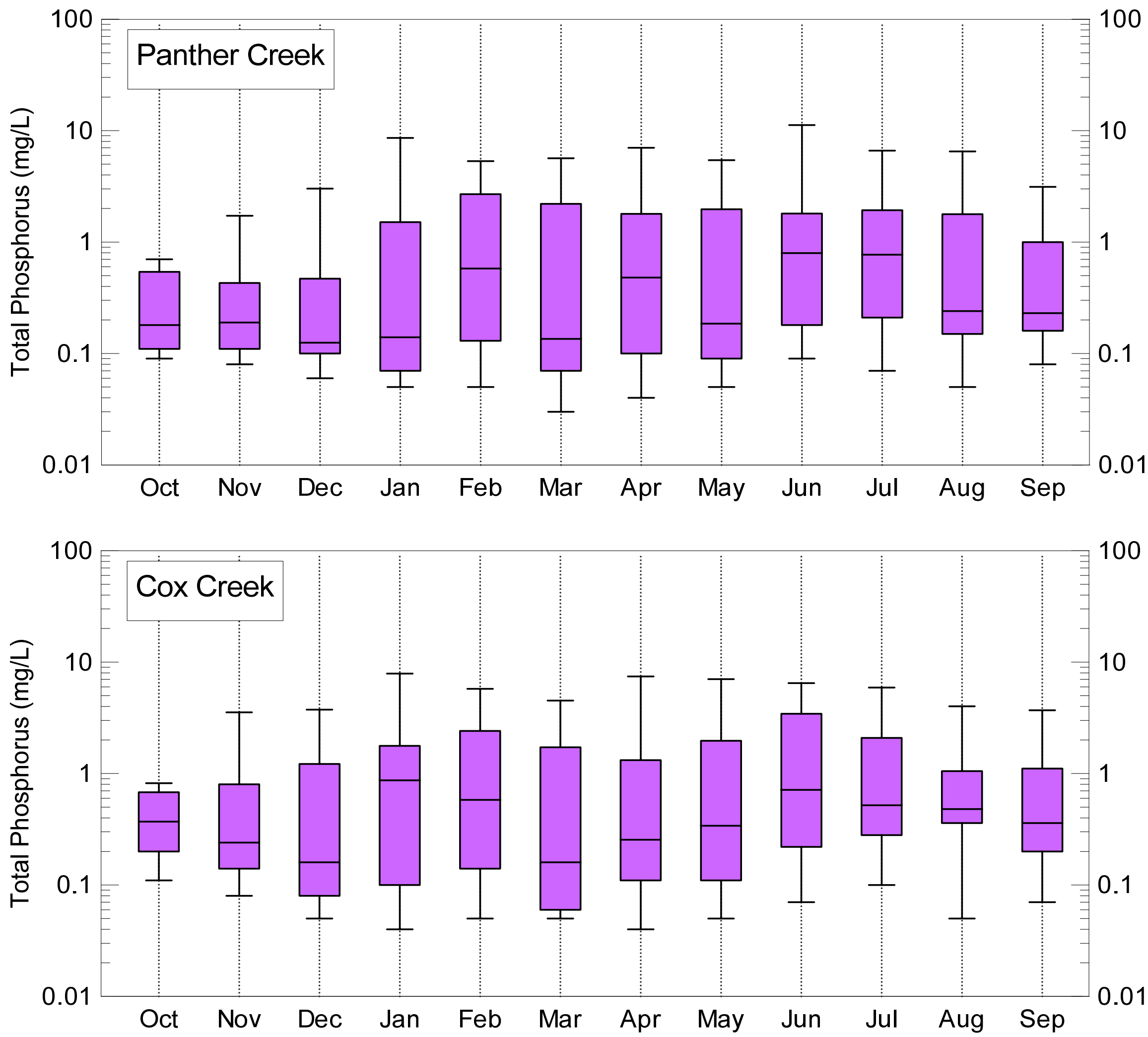

Figure 3-14. Monthly variation in total phosphorus concentrations at Sangamon sites

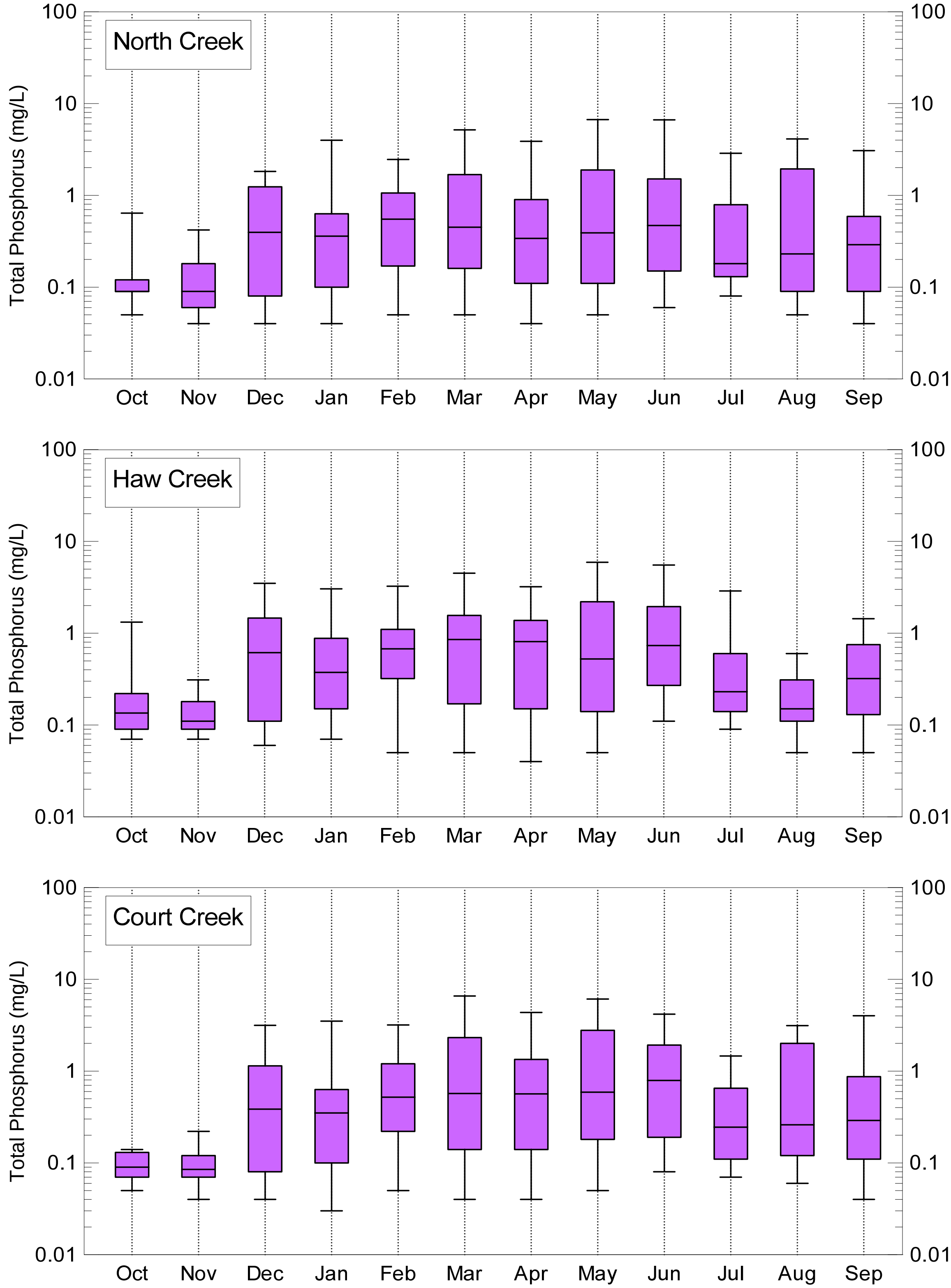

Figure 3-15. Monthly variation in total phosphorus concentrations at Spoon sites

Table 3-11. Monthly distribution of phosphorus samples collected by site

Site Number	10	11	12	1	2	3	4	5	6	7	8	9	Total Number of Samples
1	10	18	13	26	31	19	30	22	30	31	22	23	275
2	11	18	14	23	31	18	29	22	30	34	21	25	276
3	9	13	24	30	37	45	42	51	47	21	19	27	365
4	12	13	26	30	36	42	44	52	50	24	13	29	371
5	9	12	22	30	38	44	44	49	49	22	23	27	369

In this chapter the challenge of monitoring small streams to compute sediment and phosphorus loads was described. Due to the flashiness of these stream systems, the potential seasonal variation of pollutant concentrations, and the length of this data set, multiple linear regression models will be tested as the algorithm of choice for load computations at these sites. Due to the potential relationship between sediment and phosphorus and the far greater number of sediment samples than TP samples, a load calculation method that utilizes this relationship is explored in the next chapter.

4. Regression Model Development for Sediment and Phosphorus Concentrations

The previous chapter provided descriptions of four types of load algorithms commonly used to determine sediment and nutrient loads. The objective of this chapter is to develop several different multiple linear regression models and determine the best model for estimating suspended sediment and total phosphorus concentrations.

4.1 Methods and Materials

4.1.1 Data Exploration

Prior to development of regression models, some initial data exploration was performed by evaluating the relationships between suspended sediment concentration (SSC), total phosphorus (TP) concentration, and streamflow. An instantaneous discharge at the time of collection was computed for each SSC and TP sample by linear interpolation of the 15-minute streamflow record.

Scatter plots of streamflow versus SSC for each site are provided in Figure 4-1, and scatter plots of streamflow versus TP concentration for each site are provided in Figure 4-2. These figures clearly indicate a direct relationship between flow and concentration, although perhaps best described by three distinct equations for low, mid and high flows. Simon et al. (2004) attributed the flattening of the sediment transport curve at high flows to the lower silt-clay contributions to suspended sediment concentrations during high flows. Another important characteristic of these plots is the large variation in concentrations for a given flow. For example, during a flow of 10 cfs at North Creek (site #3), the expected sediment concentration could be 2-800 mg/L. At Cox Creek (site #1) during flows ranging from 0.1–10 cfs, the expected TP concentration could be 0.04–2 mg/L.

This variation in concentration for a given discharge can be partially explained by the hysteresis effect described previously in this study. On the rising limb of the hydrograph, concentrations are significantly higher than at the same flows during the receding limb. Due to the rapid changes in flow and concentrations at these study sites, this behavior can be difficult to see when looking at sample concentrations plotted with a hydrograph (Figure 4-3). Presenting concentration as a function of instantaneous discharge for a given storm event more clearly illustrates this phenomenon (Figure 4-4). Prior to the runoff event on 7/21/2008, Panther Creek was flowing at a rate of approximately 9 cfs. The daily sediment sample, collected at noon, had a concentration of 36 mg/L. Eight hours later the stream began to rise and within 30 minutes streamflow had increased to 38 cfs and sediment concentration was greater than 3,200 mg/L. Concentrations continued to increase, peaking 15 minutes later (20:45) at a concentration exceeding 9,800 mg/L (Q=179 cfs) before then decreasing at a slower rate. The peak flow during this runoff event was 586 cfs at 21:30, 45 minutes following the peak of the sediment chemograph; the concentration at the time of peak discharge was approximately 5,800 mg/L. As shown in Figure 4-4, the samples at 20:45 and 23:00 were collected at similar discharges, yet their concentrations differ by approximately 7,000 mg/L. The clockwise hysteresis exhibited during this storm event is prevalent at all study sites, although the rapidity of changes in streamflow and concentration are greater at Panther and Cox Creeks than at the larger sites in the Spoon watershed. This behavior is seen in both sediment and phosphorus data, but the TP sample coverage during runoff events is much more limited.

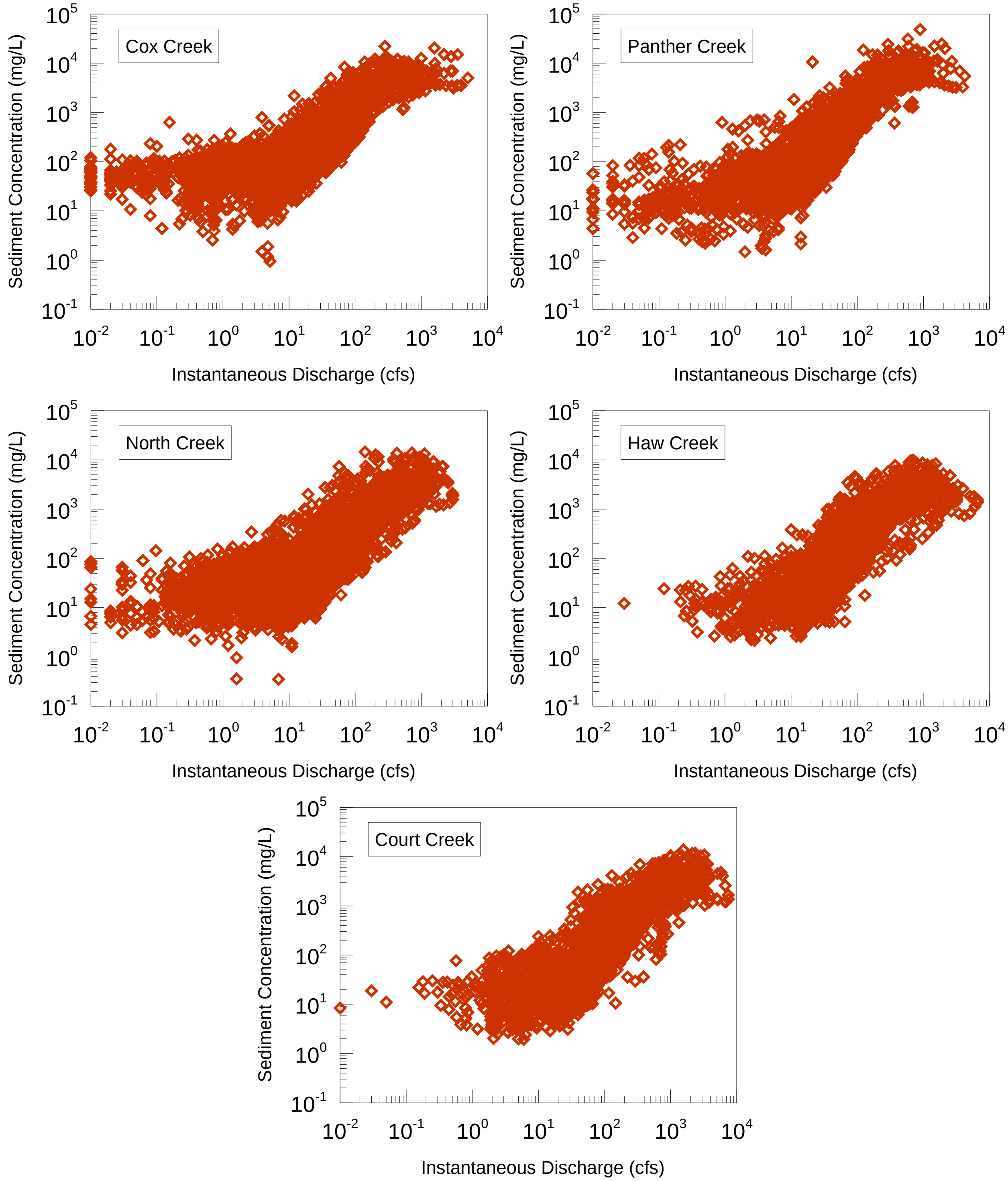

Figure 4-1. Relationship between instantaneous streamflow and suspended sediment concentration at study sites

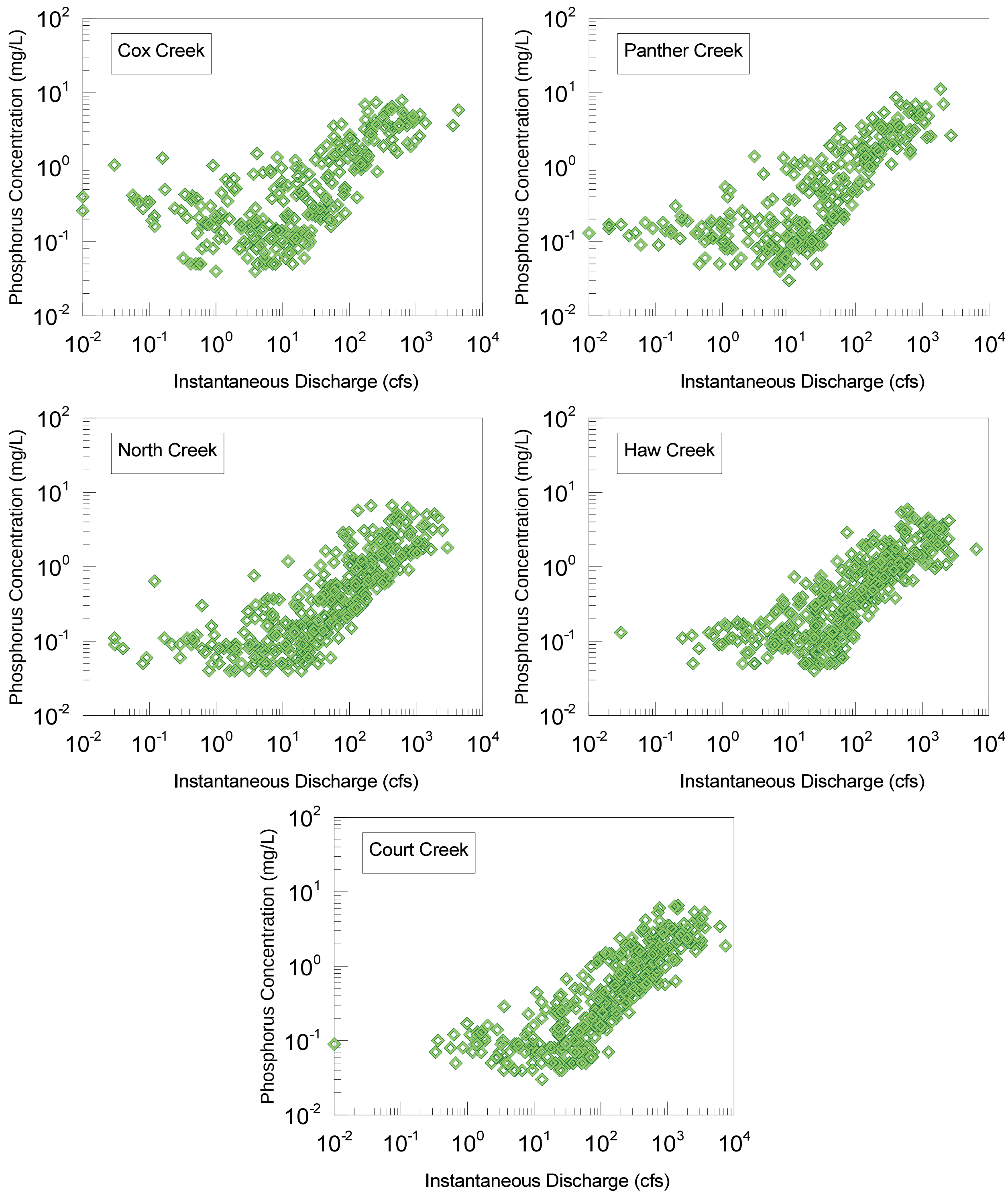

Figure 4-2. Relationship between instantaneous streamflow and total phosphorus concentration at study sites

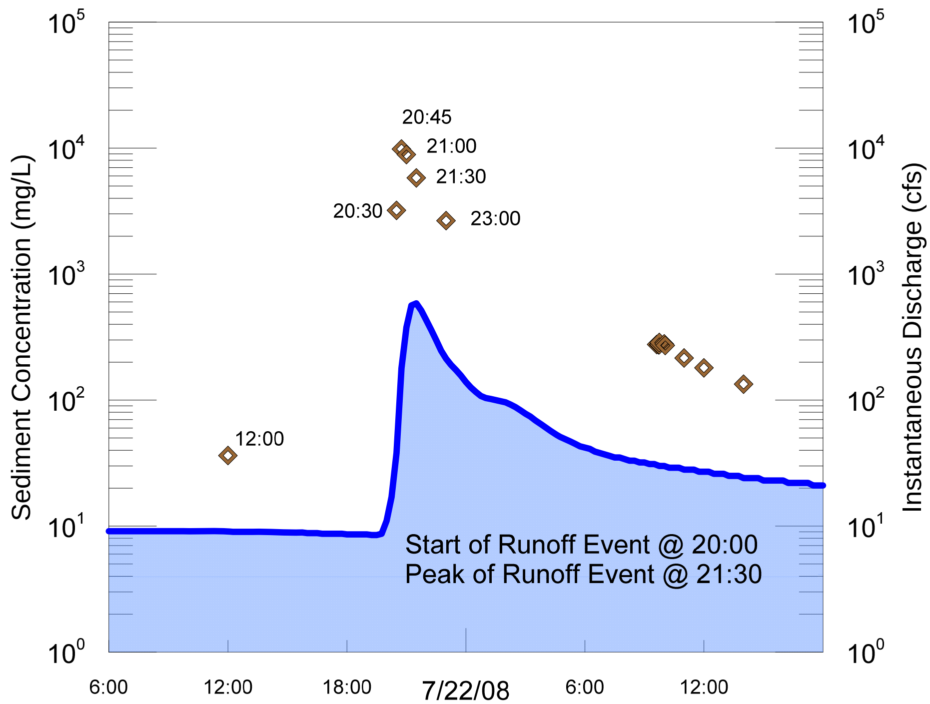

Figure 4-3. Suspended sediment concentrations during runoff event at Panther Creek, 7/21-7/22/2008

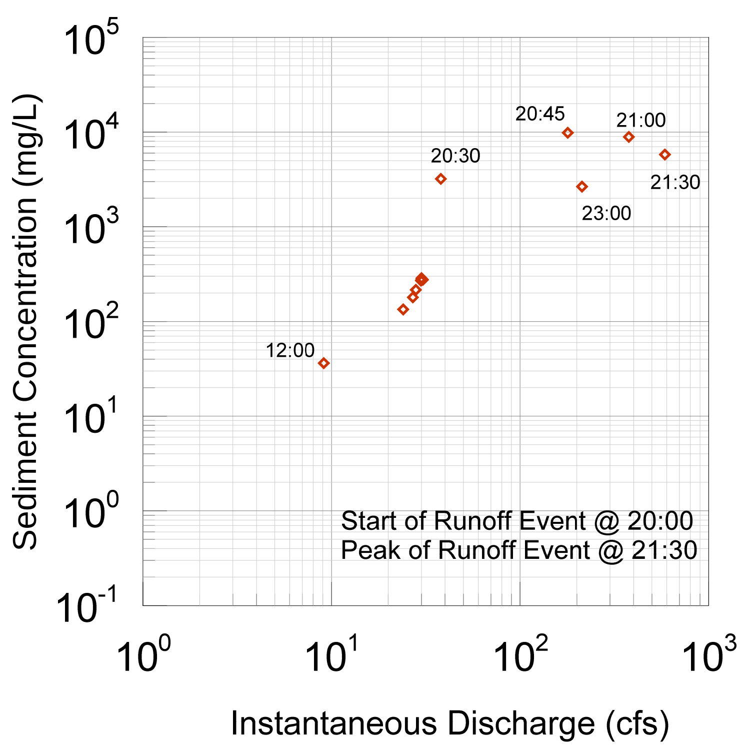

Figure 4-4. Relationship between instantaneous streamflow and suspended sediment concentrations during runoff event at Panther Creek, 7/21-7/22/2008

During routine site visits when TP is scheduled for collection, two depth-integrated samples are collected from the same location typically one immediately following the other, with one sample submitted for SSC analysis and the other for TP analysis. As a result, sediment and TP samples were rarely collected at the exact same time. In order to match up pairs of TP and SSC samples for further analyses, a small time window of 30 minutes was established during which time samples would be considered coincident. Constraining the time window to 30 minutes resulted in a small reduction (between 11 and 15%) in the overall number of TP observations available for correlation analysis (Table 4-1). The median time between paired observations was one minute for all five sites.

Table 4-1. Summary of SSC and TP pairings

Site Number	SiteName	Number of SSC Samples	Number of TP Samples	Number of SSC/TP Paired Observations	Percentage of Observations Paired
1	Cox Creek	2910	275	244	89%
2	Panther Creek	3658	276	247	89%
3	North Creek	4069	365	313	86%
4	Haw Creek	3984	371	312	84%
5	Court Creek	3258	369	314	85%

The paired samples are plotted for each site in Figure 4-5 and a direct relationship between TP and SSC is clearly evident. An equation describing log-transformed TP as a function of log-transformed SSC was fit for each site. These equation coefficients are provided in Table 4-2. Additionally all pairs of sediment and total phosphorus concentrations were ranked by increasing SSC. To measure the strength of association between the two constituents, Kendall's tau correlation coefficient was computed for each site.

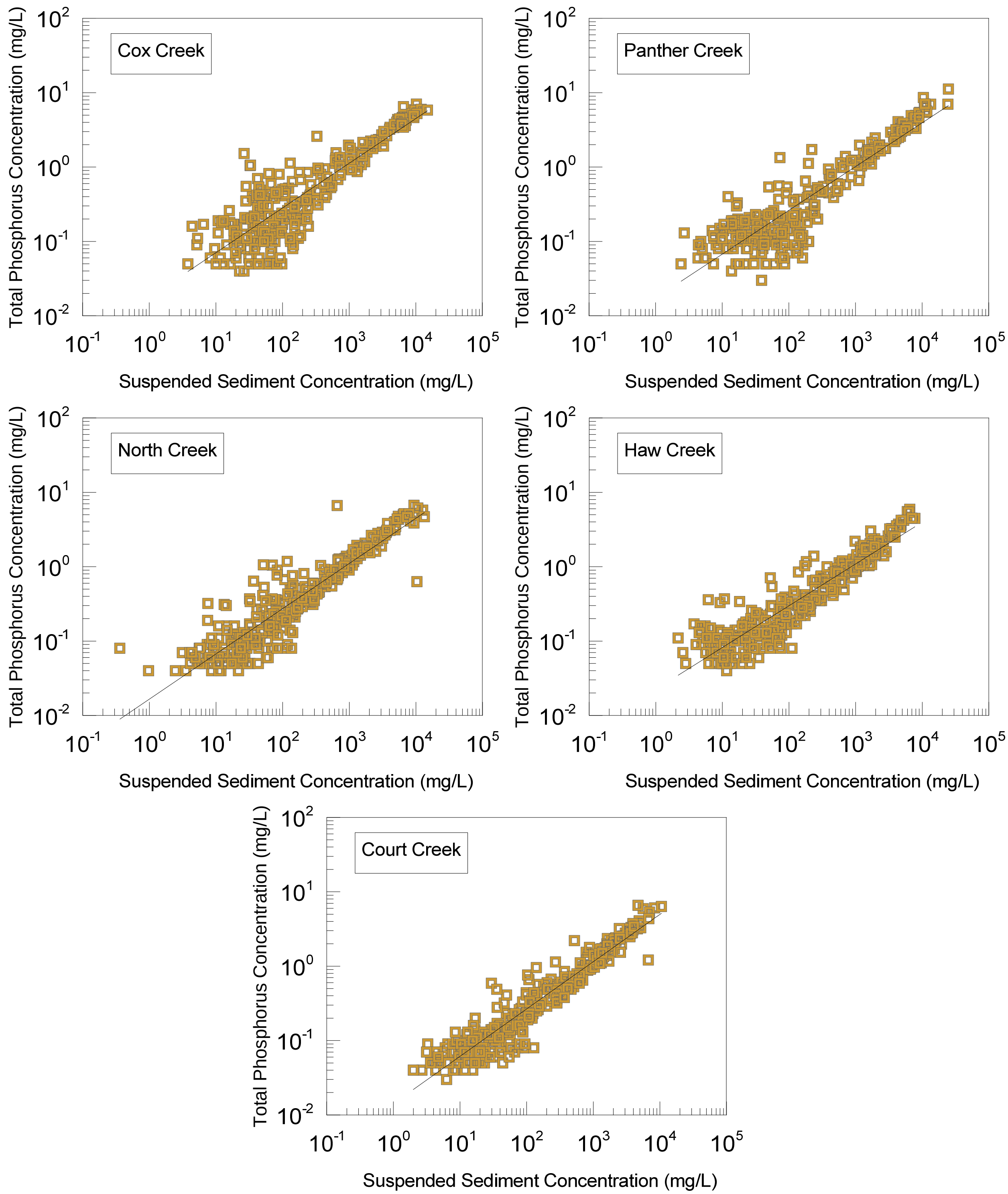

Figure 4-5. Relationship between suspended sediment and phosphorus concentrations at study sites

Table 4-2. Best-fit line coefficients and correlation statistics for SSC and TP relationships shown in Figure 4-5

Site Number	SiteName	Intercept	Slope	R^2	Tau
1	Cox Creek	-4.0335	0.6008	0.76	0.64
2	Panther Creek	-4.0520	0.5886	0.82	0.68
3	North Creek	-4.1000	0.6080	0.85	0.77
4	Haw Creek	-3.7959	0.5618	0.86	0.77
5	Court Creek	-4.2550	0.6360	0.91	0.83

The best-fit line coefficients reveal extremely similar relationships between SSC and TP among the five study sites, although the strength of the correlation increases with drainage area. Greater variability at low concentrations is seen at the three sites with the smallest drainage areas.

4.1.2 Regression Model Development

Initial data exploration supports the need for a regression model that includes terms for discharge and hydrographic position. In an effort to address further unexplained variances in concentration, explanatory variables for seasonality and trends will also be explored. Based on the results of the correlation analysis, the use of SSC as a predictor variable in TP regression models is clearly warranted.

A matrix of the explanatory variables included in each model evaluated for predicting sediment and total phosphorus concentrations at each site is provided in Table 4-3. Models 1-9 are comprised of various combinations of explanatory variables for discharge, seasonality, and long-term time trends and were selected for evaluation because these models are used by the USGS LOADEST program and allow for comparisons to be made to assess the importance of including and/or removing individual predictor variables. It should be noted that Model 9 is

Cohn's 7-parameter regression (Equation 3.3). Two additional models (10 and 11) were also evaluated because they include modifications suggested by Wang and Linker (2008) to improve prediction of the hysteresis effect through incorporation of terms to represent rates of change in flow. Based on correlations confirmed in Section 4.1 as well as the fact that sediment samples were collected much more frequently than TP samples, a 12^{th} regression model for TP was developed that included a term for instantaneous suspended sediment concentration. To investigate the potential over-parameterization of Model 12, regression models for TP based solely on SSC (Model 13) as well as SSC and discharge (Model 14) were also evaluated.

Table 4-3. Parameters in regression models evaluated

Model	intercept	lnQ	$(\ln Q)^2$	sin(2πT)	cos(2πT)	T	T^2	ln(dQ)	1/(lnQ)	lnSSC
1	X	X								
2	X	X	X							
3	X	X				X				
4	X	X		X	X					
5	X	X	X			X				
6	X	X	X	X	X					
7	X	X		X	X	X				
8	X	X	X	X	X	X				
9	X	X	X	X	X	X	X			
10	X	X	X	X	X	X	X	X		
11	X	X	X	X	X	X	X	X	X	
12*	X	X	X	X	X	X	X	X		X
13*	X									X
14*	X	X								X

*Models developed for TP prediction only.

To prepare for model development, calibration files consisting of the following five fields were created for each site: observed sample concentration, Q, T, (2πT), and dQ. The discharge variable (Q) was the instantaneous discharge at the time of sample collection. Because several of the study sites experience periods of zero flow, a small constant (0.01) was added to discharge prior to log-transformation. Additionally discharge was reduced by a centering constant computed for each site to eliminate collinearity between linear and quadratic terms (Cohn et al., 1992). The time of sample collection (T) was expressed in units of decimal years, and the trend variables (T, T^2) were also adjusted by a centering constant while the seasonality variables were not. The rate of change in flow at the time of sample collection (dQ) was computed as the difference in flow during the 15-minute interval bounding the time of sample collection. After log-transformation, this can be expressed as $\ln Q_i - \ln Q_{i-1}$ or its equivalent, $\ln(Q_i/Q_{i-1})$.

Calibration files for TP models which include SSC as a predictor variable (Models 12-14) will be a smaller calibration data set because not every TP sample has a coincident SSC sample (Table 4-1). In order to evaluate any reductions in model performance due to the smaller TP calibration dataset, Models 1-11 were evaluated for the full TP calibration data set (e.g. n=275 for site #1), while the subset of TP samples that had a paired SSC sample (e.g. n=244 for site #1) were used for evaluating Models 1-14.

Development of calibration data files and all subsequent regression analysis was performed using MATLAB.

4.1.3 Regression Model Evaluation

Model performance was evaluated through a combination of visual diagnostics and performance statistics. For each model the following four plots were created: (1) observed vs. predicted concentrations, (2) observed concentrations (log-transformed) vs. residuals, (3) observed concentration (original units) vs. residuals, (4) predicted concentrations vs. residuals. The statistics evaluated include Mean Square Error (MSE), the adjusted R^2 value, nested F statistic, and the Akaike's Information Criterion (AIC).

The adjusted R^2 value is simply the traditional R^2 value weighted by the ratio of total degrees of freedom to error degrees of freedom. This is a less sensitive measure of model quality for SSC due to the sheer number of data points compared to the model variables. This metric is useful for comparing the proportion of variation explained by selected models, but because R^2 increases as the number of variables increase it is not recommended to use solely for selection of the best model.

To determine whether the addition of additional model variables led to improvements in the model, the nested F test was performed. The Nested F test (Helsel and Hirsch, 2002) compares the decreased error to the loss of degrees of freedom in the more complex model and is defined as

$$F = \frac{(SSE_s - SSE_c)/(df_s - df_c)}{(SSE_c / df_c)} \qquad \textbf{(4.1)}$$

where s refers to the simpler and c the more complex model. If the F-ratio is small then the additional variables offer little improvement to the model.

AIC (Burnham and Anderson, 2002) was also used for model selection because it can be used to compare non-nested models (i.e. Model 1 vs. Model 13).

$$AIC = n \times \log(SSE / n) + 2(p + 1) \qquad \textbf{(4.2)}$$

where n is the number of observations and p is the number of model parameters. Lower AIC values indicate a better model. The magnitude of AIC is not important, but rather the relative AIC value as compared to other models based on the same dataset is the important metric.

4.2 Results and Discussion

The use of instantaneous discharge as a predictor variable is a key difference in this study. The overwhelming majority of studies found in the literature use mean daily discharge in regression models for load estimation. Even two recent studies exploring different sampling strategies for load estimation on small streams used regression models developed from mean daily discharge (Robertson, 2003; Robertson and Roerisch, 1999). While Wang and Linker (2008) added regression model terms for rate of change in flow, they were using mean daily discharge data and they theorized that the impact of these model terms would be more pronounced when utilized with data of a smaller time-step.

Using 15-minute records of streamflow from small, flashy streams presented additional challenges in regression model development. The rate of change in flow term, ln(dQ), led to predictions of unrealistically high concentrations during transitions from zero or very low flow to sudden rises in the stream. To address this issue, an upper-bound was placed on the value of this term in the regression models. To determine the upper limit for this term, the entire series of log-transformed rates of change in streamflow between 15-minute intervals was plotted for all stations for the 10-year period of study, and this information was used to explore the distribution of rates of change in streamflow and whether a few selected values were, in fact, extreme. An upper-limit of 3 was selected for the log-transformed rate of change of flow, which is equivalent to a twenty-fold increase of flow within 15 minutes (i.e. 2 cfs to 40 cfs).

4.2.1 Suspended Sediment Models

The model coefficients and their standard errors were evaluated for all 11 SSC models at each site. The linear and quadratic discharge terms were statistically significant at the 0.01 probability level in all sediment models at all study sites. Both seasonality terms were also significant at the 0.01 level in all models at all sites, with the exception of Models 10 and 11 at Cox Creek, where the sine term was significant at the 0.05 level. The quadratic time trend variable was significant at the 0.01 level in all models at all sites, however the significance of the linear time trend parameter varied between the sites. At Panther Creek and Court Creek the linear time trend variable was significant at the 0.01 level in all models, and at Cox Creek this term was significant at the 0.01 level in all models except Model 5 (p-value=0.05). At Haw Creek the time trend variable was significant at the 0.01 level in Models 3 and 5, at the 0.05 level in Model 7, and at the 0.10 level in Models 8-11. At North Creek this term was significant at the 0.01 level in Model 3 and at the 0.05 level in Model 5 but was not statistically significant in the remaining models. The variable describing the rate of change in flow, ln(dQ), was statistically significant at the 0.01 level at all sites. The inverse flow variable (1/lnQ) however was found to not be statistically significant at the three Spoon watershed sites but was significant at Cox Creek (p-value=0.01) and Panther Creek (p-value=0.05).

Plots of MSE, adjusted R^2, and AIC values for all sediment models at the five study sites are presented in Figure 4-6. While the variation in model performance was greater at the three smallest study sites and more consistent at the two larger sites in the Spoon watershed, Haw Creek and Court Creek, the overall model performance was consistent across the study sites. Models 10 and 11 appear to be the best choices for SSC models, followed closely by Model 9. Evaluating the nested F statistic for Panther Creek revealed that the addition of an 8^{th} variable (Model 10) was a significant improvement over Model 9, but the improvement in model performance from Model 11 to Model 10 was at a much lower level of significance.

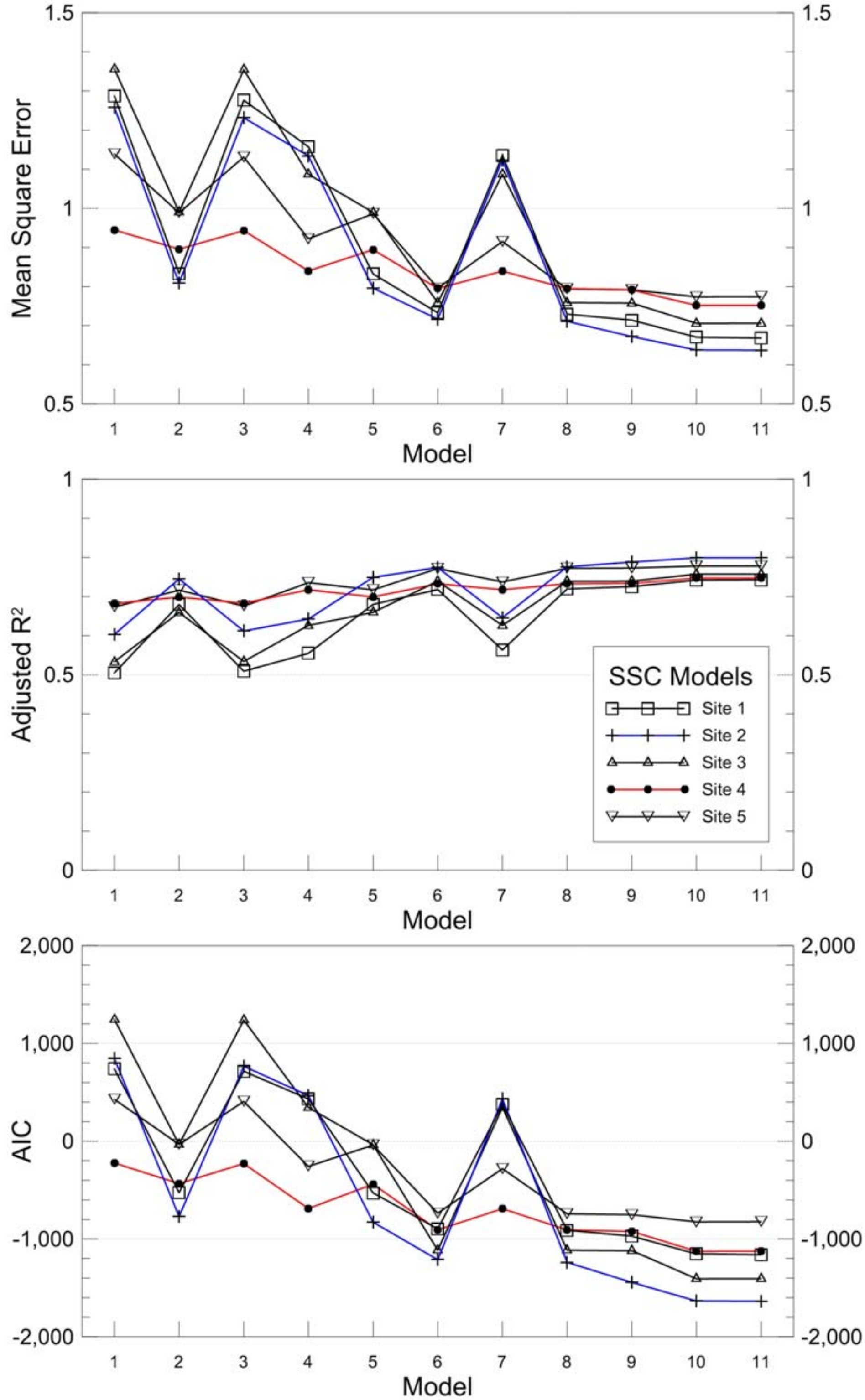

Figure 4-6. Performance statistics for SSC models developed for study sites. Note that site IDs were assigned in order of increasing drainage area.

Based on the plots of observed vs. predicted concentrations, Models 1, 3, 4 and 7 (which do not contain the quadratic discharge term) tend to fit the low and mid-range concentrations well, but underestimate the highest sediment concentrations. All remaining models tend to overestimate very high sediment concentrations.

4.2.2 Total Phosphorus Models

The model coefficients and their standard errors were evaluated for all 11 TP models developed using the full calibration dataset. The linear and quadratic discharge terms were statistically significant at the 0.01 probability level in all TP models at all study sites. Both seasonality terms were also significant at the 0.01 level in all models for Cox, Panther, Haw and Court Creeks, with the exception of Model 7 at Cox Creek (p=0.05). At North Creek the cosine term was significant at varying levels across all models: Model 9 (p=0.01); Models 6, 8, 10, and 11 (p=0.05); and Models 4 and 7 (p=0.10). The significance of the linear and quadratic time trend parameters varied greatly across study sites. None of the time trend parameters were significant at Cox Creek. At Panther Creek the time trend variables were significant at the 0.05 level in Model 5 and at the 0.10 level in Models 5 and 8-11. The linear time trend variable was not significant at any of the three Spoon watershed sites, however the quadratic time trend variables was found to be significant at the 0.01 probability level at these sites. The variable describing the rate of change in flow, ln(dQ), was statistically significant at the 0.01 level at Cox Creek and all Spoon watershed sites and at the 0.05 level at Panther Creek. The inverse flow variable (1/lnQ) was found to not be statistically significant at any of the study sites.

The model coefficients and their standard errors were also evaluated for the 14 TP models developed using the smaller calibration dataset of TP samples with coincident SSC samples. In general the significance of these model parameters was consistent with the significance levels determined from the full calibration data set, with only occasional minor differences in the level of significance. The instantaneous SSC term was found to be statistically significant at the 0.01 level in Models 12-14 at all study sites. With the addition of SSC as a predictor variable in Model 12, the rate of change in flow term was no longer statistically significant at Panther, Cox, and Court Creeks.

Plots of MSE, adjusted R^2, and AIC values for the 14 TP models based on the smaller calibration dataset at the five study sites are presented in Figure 4-7. Similar to the SSC models the variation in model performance was greater at the three smallest study sites than at Haw Creek and Court Creek, and the pattern of model performance was consistent across the study sites. Model 12 appears to be the best choice for TP models, followed by Models 13 and 14.

Based on the plots of observed vs. predicted concentrations, Models 1, 3, 4 and 7 underestimated the highest TP concentrations as well. The remaining models seemed to fit the entire range of TP concentrations well.

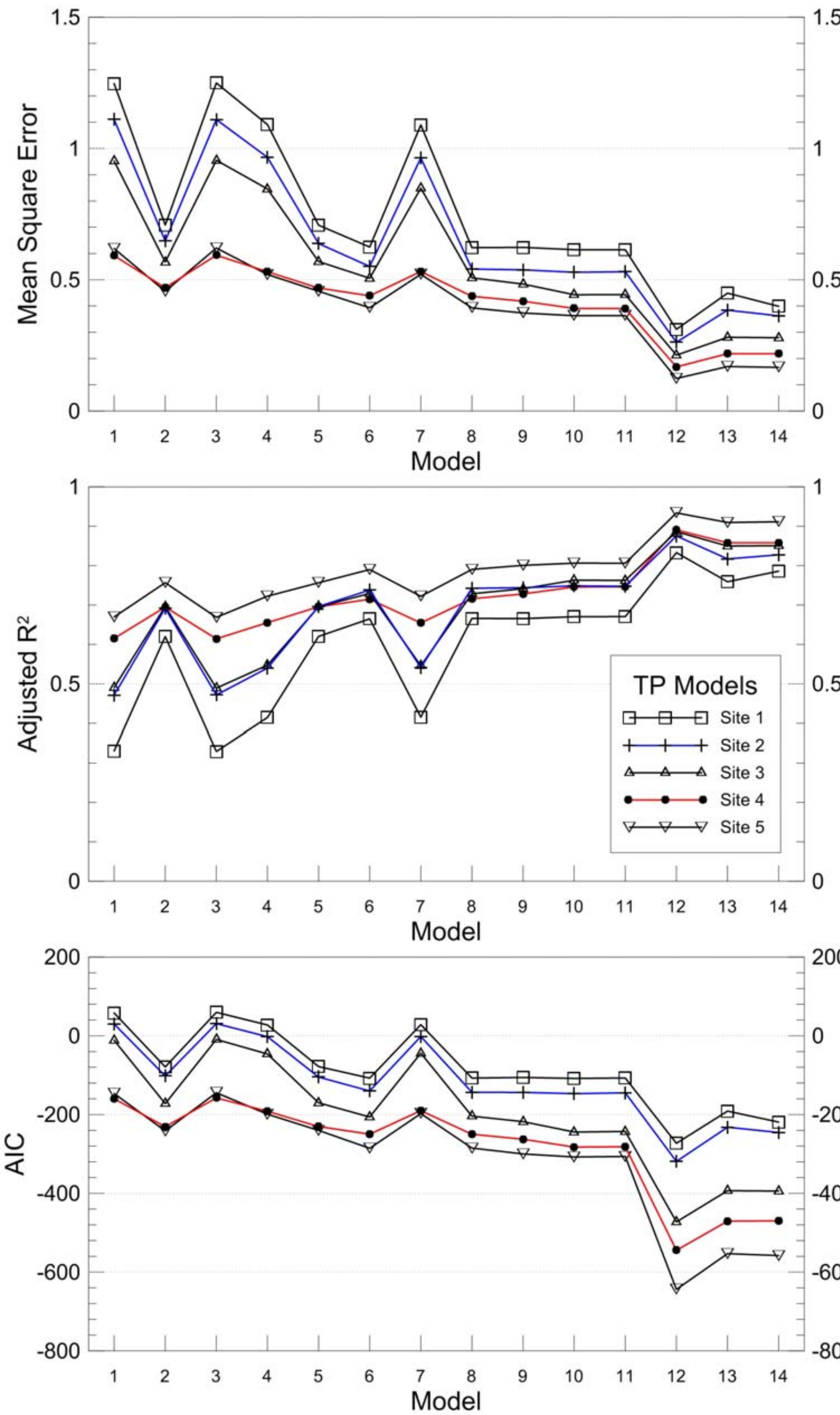

Figure 4-7. Performance statistics for TP models developed for study sites

4.3 Summary and Conclusions

Eleven log-linear regression models were developed for sediment concentration prediction for each study site. Based on various performance statistics, Model 10 was selected as the best model but Model 9 will also be evaluated in the next chapter to allow for comparisons to Cohn's 7-parameter regression equation. According to AIC values, Model 11 was rated as a slightly better model than 10 for Panther Creek and Cox Creek. However the nested F statistic did not strongly support this, and to simplify calculation of models across all sites Model 10 was chosen for all sites.

Fourteen log-linear regression models were developed for total phosphorus concentration prediction for each study site. Based on various performance statistics, Model 12 (9-parameters) was selected as the best model but Model 13 (2-parameters) will also be evaluated in the next chapter to allow for comparisons to a much simpler equation based solely on TP's relationship to SSC. For additional comparisons, Models 10 and 9 will also be used with the full TP calibration dataset for future load calculations. According to AIC values, when not using sediment as a predictor variable Model 10 was clearly the best model at all sites for TP prediction.

To facilitate model application across all sites for both constituents, all terms were retained in the selected models even if the model parameters were not significant at one of the study sites. It is expected that inclusion of the insignificant terms and the estimation of the additional parameter(s) will cause only a small proportional decrease in the degrees of freedom in the regression because of the large number of observed samples available. Additionally, often the sine term was significant but the cosine term was not, and it was not appropriate to remove one without the other.

The selected sediment models over-estimated concentrations at the highest flow rates. This is not unexpected given the change in slope of the discharge-concentration relationship evident at the five study sites (Figure 4-1). While these high predicted concentrations are a concern, the next chapter will explore the use of error correction techniques to use the information contained in the model residuals to improve these concentration estimates and subsequent load calculations.

5. COMPARISON OF LOAD CALCULATION TECHNIQUES

In order to determine the best approach for computing loads at the five study sites given the large number of samples available, the objective of this chapter is to evaluate load calculations using the regression equations selected in the previous chapter as compared to load calculations performed using simple linear interpolation of observed samples, the composite method, and a modified composite method.

Even intensive monitoring programs with dedicated storm sampling have periods of missing data or under-sampled storm events. An important consideration when selecting a load calculation technique is how these periods between observed samples are estimated. Given the high number of observed samples, frequency of event sampling, and availability of continuous streamflow records, the best approach for computing loads is estimating continuous (15-minute) records of concentration, also known as chemographs, to subsequently multiply by the streamflow record as opposed to other methods of directly estimating loads (Porterfield, 1972).

5.1 METHODS AND MATERIALS

5.1.1 Development of Continuous Records of Concentration

Several scripts were developed in MATLAB to create continuous chemographs using four different estimation methods: linear interpolation (LI); regression models (RM#, where # is the model selected); regression models with error correction using the composite method (RM#_CM); and regression models with error correction using the rectangular proportional method (RM#_RP).

Linear interpolation between observed samples is often used to estimate missing record when sample coverage is extensive and the proportion of missing record is small (Johnes, 2007;

Royer et al., 2006; Robertson, 2003). Linear interpolation is computationally simple but does not take into consideration flow, seasonality, hydrographic position, or any other potentially explanatory factor. However, this method may be sufficient with a large data set of observations collected throughout a variety of flow conditions and seasons. Chemographs were developed for each site by linearly interpolating between each observed sample a concentration on a fixed 15-minute time step for the 10-year study period (WY 2000-2009). A small portion (16 hours) of a chemograph developed by linear interpolation is presented in Figure 5-1, along with the corresponding observed samples.

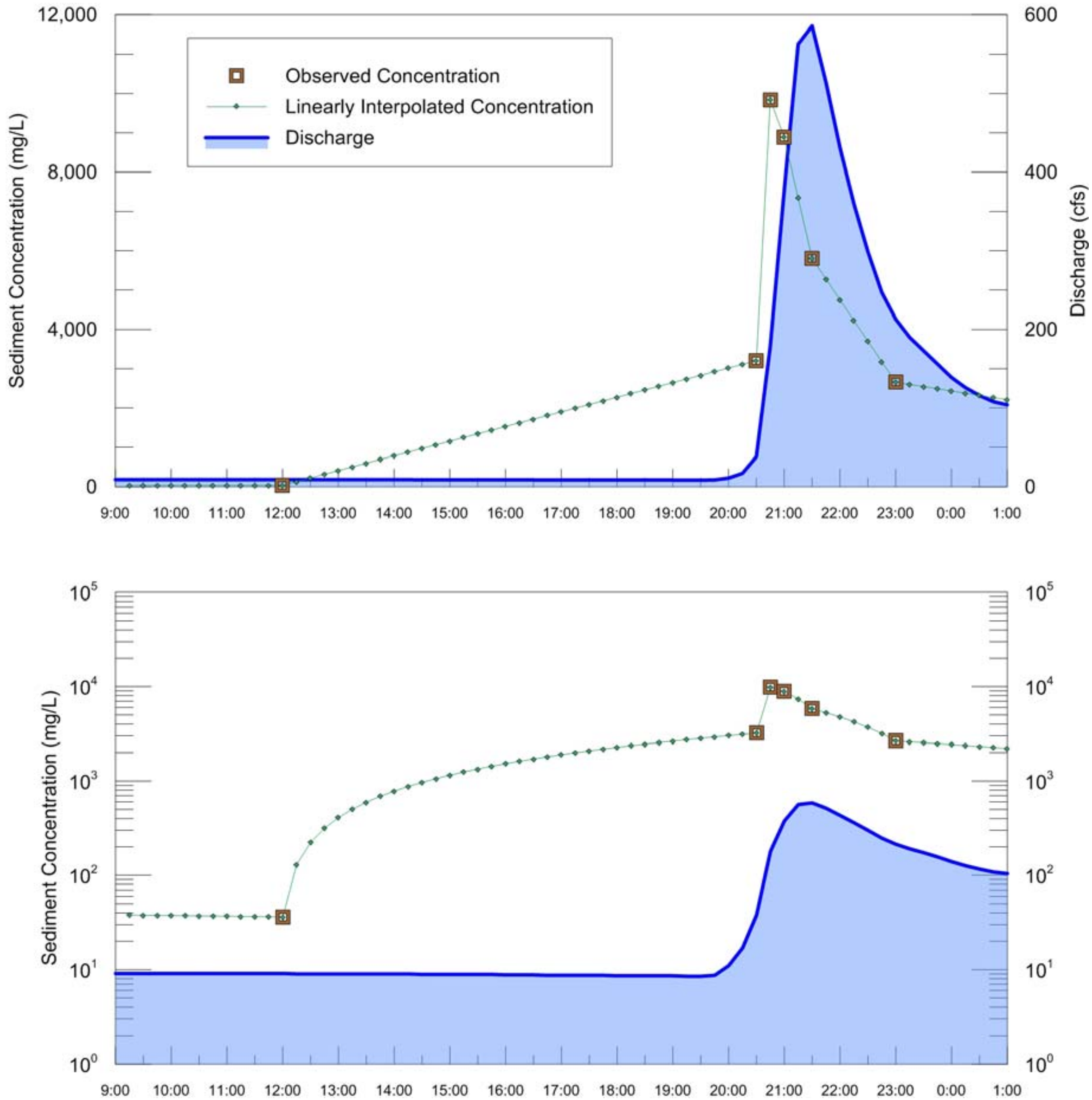

Figure 5-1. Example of linear interpolation between observed samples to construct 15-minute chemograph presented on both a linear scale (top) and logarithmic scale (bottom)

To apply the selected regression models (9, 10, 12 and 13), estimation files were created for each site which contained measures of each explanatory variable (Table 4-3) computed on a 15-minute time step for the 10-year study period (WY 2000-2009). Because TP Models 12 and 13 include SSC as a predictor variable, the estimation files for these models will require a 15-

minute record of SSC. The best method of concentration estimation for SSC as determined subsequently in this chapter will be used in the estimation files for Models 12 and 13. To develop the chemographs utilizing regression models, each site's estimation file will be multiplied by the regression coefficients developed in the previous chapter. A chemograph developed from a regression model for the same 16-hour period is presented in Figure 5-2.

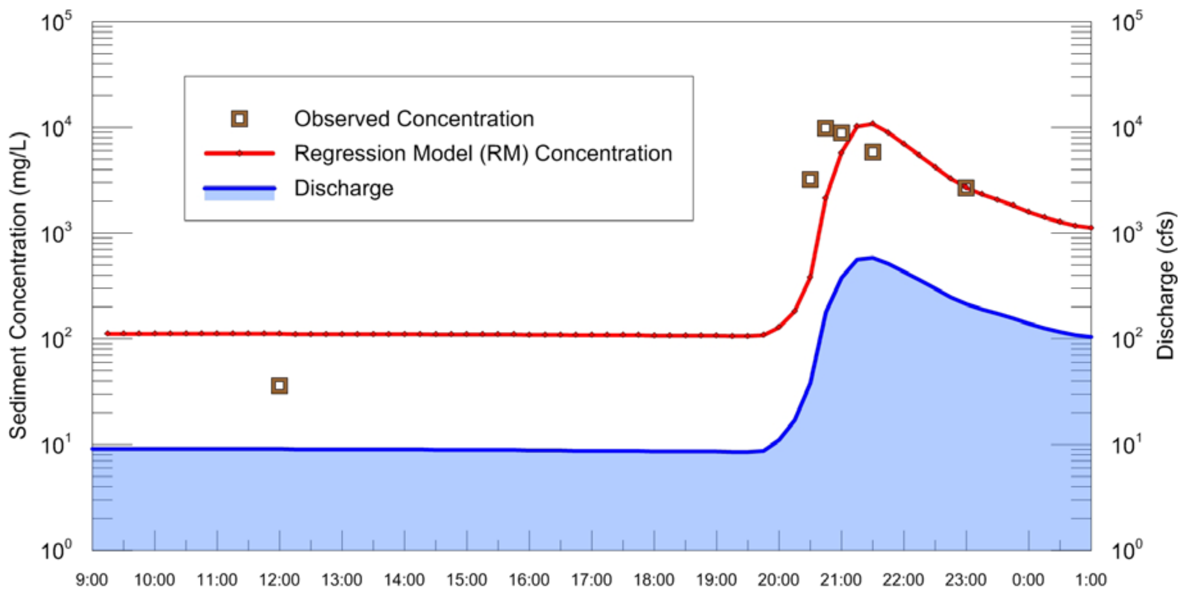

Figure 5-2. Example of a 15-minute chemograph constructed from a regression model

To apply the composite method as described by Aulenbach and Hooper (2006), a continuous record of regression model residuals was developed by linear interpolation between the residuals of observed samples. These residuals were then used to adjust the regression model predicted concentrations, which in effect sets the model predicted concentrations equal to the observed concentrations. A chemograph developed from the composite method (CM) for the same 16-hour period is presented in Figure 5-3.

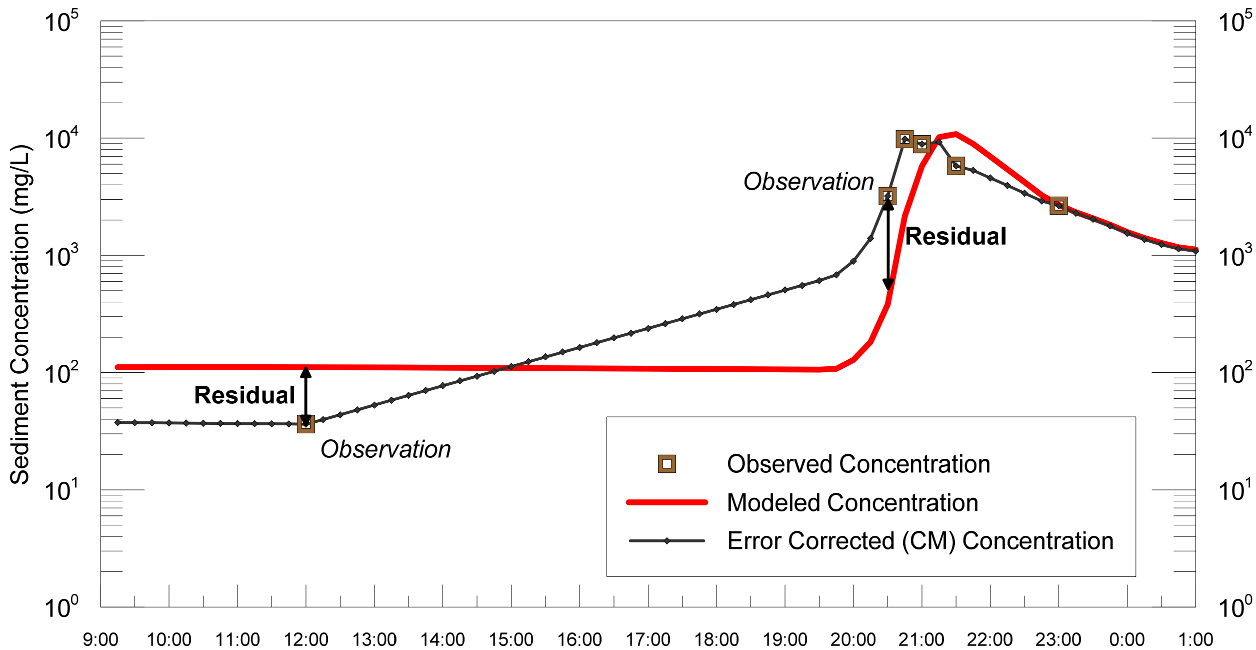

Figure 5-3. Example of a 15-minute chemograph constructed from a regression model with error correction using the composite method

The fourth method of chemograph construction evaluated was the modified composite method presented by Verma et al. (2012), who found that applying the composite method using the rectangular proportional approach to adjust model predictions was the best approach for modeling nitrate-N in the Vermilion River in Illinois. A chemograph developed from the modified composite method, utilizing rectangular proportional (RP) error correction for the same 16-hour period is presented in Figure 5-4.

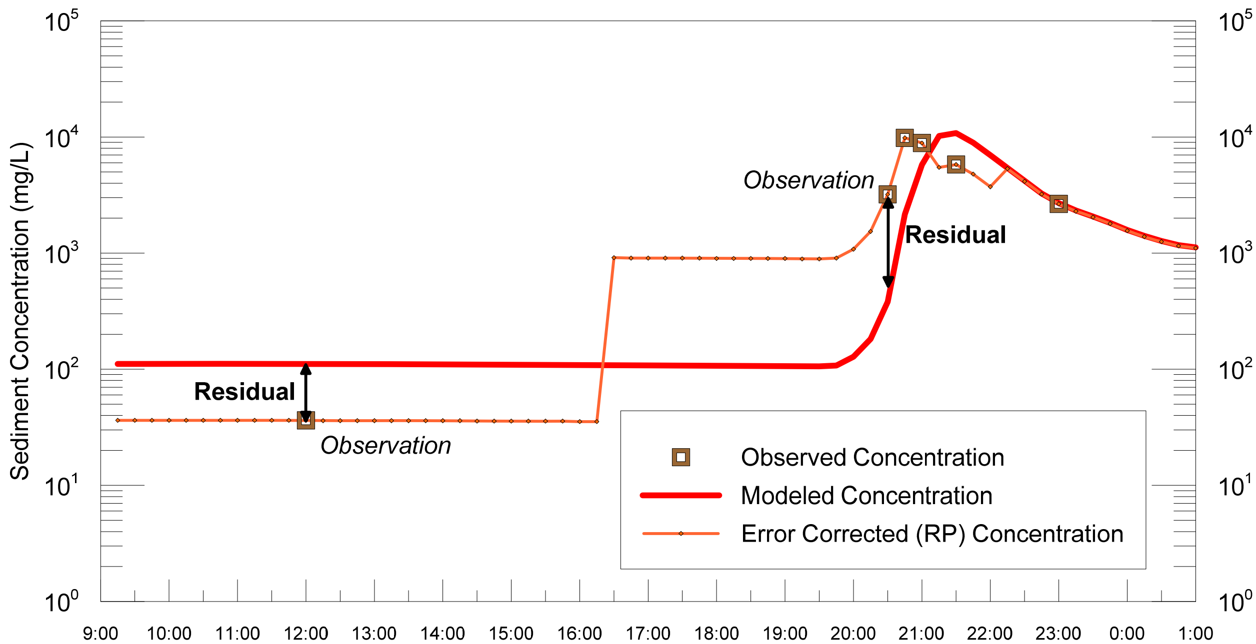

Figure 5-4. Example of a 15-minute chemograph constructed from a regression model with error correction using the modified composite method (rectangular proportional)

For each regression model selected for evaluation, both types of error correction were also evaluated at each study site. The selected models with and without error correction evaluated in this study and the notation used in subsequent figures and discussions are presented in Table 5-1.

Table 5-1. Methods of chemograph construction evaluated for each constituent

Suspended Sediment	Total Phosphorus
LI	LI
RM9	RM9
RM9_CM	RM9_CM
RM9_RP	RM9_RP
RM10	RM10
RM10_CM	RM10_CM
RM10_RP	RM10_RP
	LI_tpSSC*
	RM12
	RM12_CM
	RM12_RP
	RM13
	RM13_CM
	RM13_RP

*Note: Linear interpolation using the smaller subset of TP samples (those with a paired SSC observation) is denoted by LI_tpSSC

5.1.2 Validation of Concentration Estimation Methods

Because three of the four concentration estimation methods evaluated use observed data, traditional measures of goodness-of-fit could not be employed. Therefore, the estimation methods were validated using a second data set collected at two of the study sites, Court Creek and North Creek, between 2004 and 2007 as part of a separate Illinois State Water Survey study evaluating the bioavailability of phosphorus in an agriculturally dominated watershed (Machesky et al., 2010). This second (validation) data set was collected at the same locations on Court Creek and North Creek, utilizing the same sampling procedures, and analyzed by the same laboratories as the primary data sets in this study. The sediment and phosphorus samples in the validation data sets were primarily collected on a bi-weekly schedule, although limited event sampling was also performed. Selected summary statistics for the validation data set are provided in Table 5-2.

Table 5-2. Summary of validation data set concentrations (mg/L), April 2004-October 2007

		Court Creek	North Creek
Suspended Sediment			
	Maximum	4,045	4,417
	Median	27	22
	Minimum	3	3
	Count	132	133
Total Phosphorus			
	Maximum	5.23	3.15
	Median	0.12	0.12
	Minimum	0.04	0.04
	Count	143	143

The methods of chemograph construction were evaluated by using the validation data set to compute statistics recommended by Moriasi et al. (2007).

The Nash-Sutcliffe model efficiency was computed using the following equation.

$$NSE = 1 - \left[\frac{\sum (Y_i^{obs} - Y_i^{sim})^2}{\sum (Y_i^{obs} - Y^{mean})^2} \right] \tag{5.1}$$

The NSE coefficient ranges from -infinity to 1, with higher values generally indicating a better fit. Values less than zero indicate cases in which the concentration estimation method is a worse representation than if one were to simply use the sample mean of the observed concentrations.

Percent Bias expresses deviation of the data as a percentage. The optimal value is zero and positive values indicate the model is underestimating concentrations and negative values indicate overestimation by the model. It is computed as:

$$PBIAS = \left[\frac{\sum (Y_i^{obs} - Y_i^{sim}) \times 100}{\sum (Y_i^{obs})} \right] \tag{5.2}$$

RMSE-observations standard deviation ratio (RSR) was computed using the following equation.

$$RSR = \frac{RMSE}{STDEV_{obs}} = \left[\frac{\sqrt{\sum (Y_i^{obs} - Y_i^{sim})^2}}{\sqrt{\sum (Y_i^{obs} - Y^{mean})^2}} \right] \tag{5.3}$$

An optimal RSR value of zero indicates that there is no residual variation, so it is a perfect model.

5.1.3 Load Calculations

Because all of the chemographs were constructed in log space, the continuously simulated records of concentration were transformed back into their original units before being multiplied by the instantaneous measures of discharge and the appropriate conversion factor to produce a 15-minute record of sediment and phosphorus loading rates. The daily loads were computed by summing these 96 loading rates per day, while monthly and annual loads were computed by summing these daily loads.

5.2 Results and Discussion

To explore the performance of these chemograph construction techniques, first the difference in concentration distributions between methods was examined. Secondly the flow conditions under which these differences in predicted concentrations occurred were compared along with the resultant difference in predicted loads.

5.2.1 Comparison of Sediment Concentration Estimates

Individual chemographs and hydrographs were compared visually to determine the characteristics of the different chemograph construction techniques. Because a large flush of sediment is typically transported during a runoff event, concentrations estimated using linear interpolation (LI) between observations typically led to under-estimation of concentrations when the runoff event was not sampled or the start of sampling was delayed such that no samples were collected on the rising limb of the flow event. The LI method would also overestimate concentrations during low flow conditions preceding a high flow event if a sample was not collected immediately prior to the start of the runoff event and thus concentrations were interpolated from a preceding sample collected much, much earlier. For example, the chemograph and hydrograph in Figure 5-1 illustrate this behavior, as the SSC in the stream is estimated to have increased to greater than 1,000 mg/L more than 5 hours prior to the start of the runoff event.

In addition to reviewing time-series plots, graphs of concentration versus discharge (QC plots) were also utilized to explore the differing behavior of the chemograph construction methods. These types of plots are especially useful for comparing the differences between different types of regression models and the effects of error correction on concentration estimation. Example plots for a runoff event at Panther Creek are provided in Figure 5-5.

While difficult to discern from the time-series plot (Figure 5-5a (left)), the corresponding QC plot (Figure 5-5a (right)) illustrates a case where the LI method creates a clockwise hysteresis effect at the highest flows, but in the middle of the receding limb this method is actually predicting concentrations higher than those experienced at the same flows on the rising limb. Regression model 9 (Cohn's 7-parameter equation) does not simulate any hysteresis effect in Figure 5-5b, predicting the same concentrations on the rising and falling limbs of the hydrograph. By simply comparing the time-series plots of RM9 (Figure 5-5b (left)) and RM10 (Figure 5-5c (left)), RM10 does not appear to predict concentrations much different from RM9. However, more insight can be gained by exploring the respective QC plots (Figure 5-5b (right) and Figure 5-5c (right)). While RM10, with its addition of an 8th parameter for the rate of change of streamflow at the time of sample collection, does in fact simulate a clock-wise hysteresis effect, it does not agree very closely with the observed sample concentrations. The impact and benefit of error correction is clear in Figure 5-5d, where concentrations simulated by regression model 10 with error correction by the composite method (RM10_CM) simulated the clock-wise hysteresis effect and matched each observed sample's concentration.

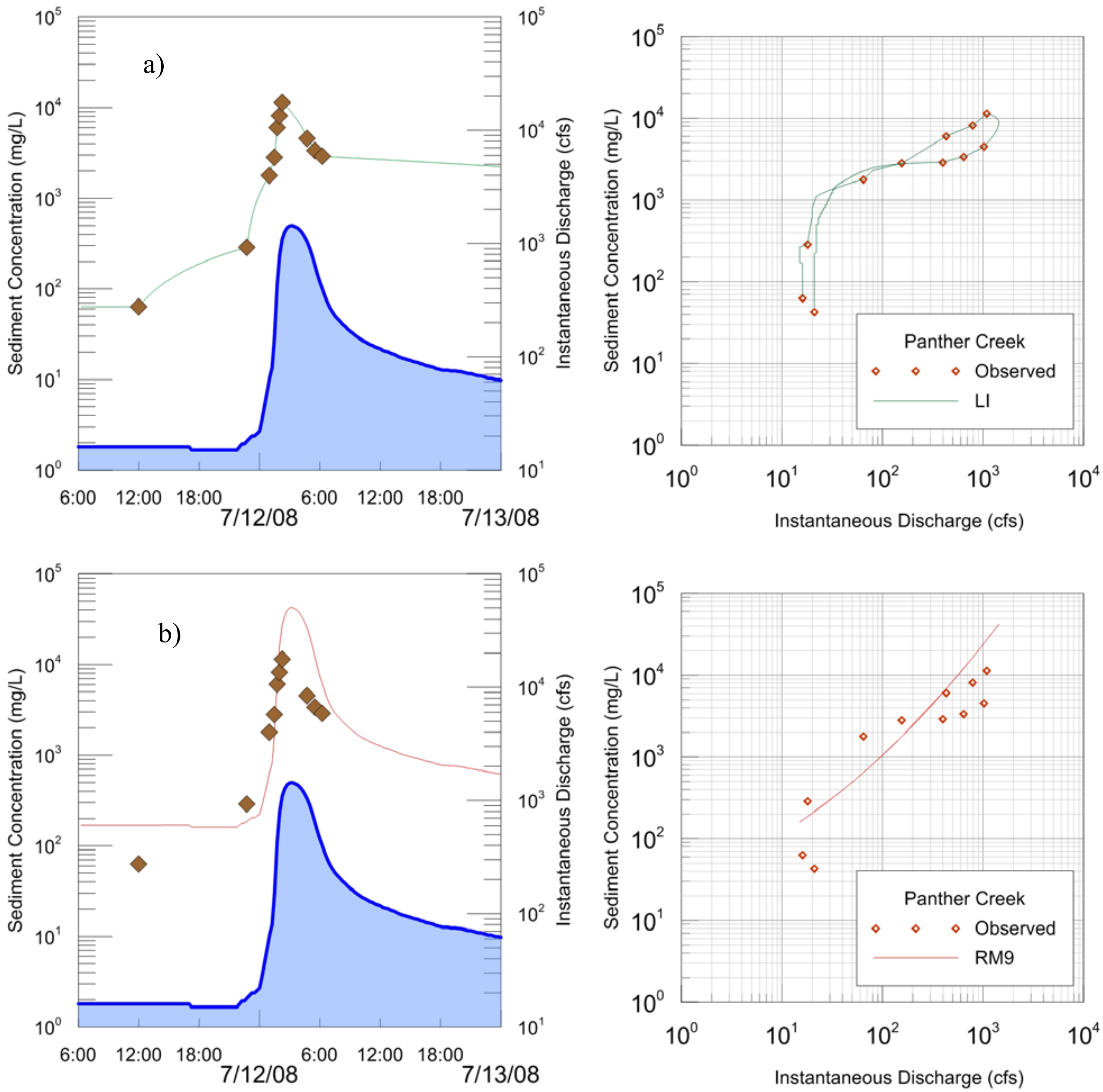

Figure 5-5. Simulated suspended sediment concentrations during runoff event plotted versus time (left) and versus instantaneous streamflow (right) at Panther Creek, 7/11-7/12/2008 using a) linear interpolation, b) regression model 9, c) regression model 10 and d) regression model 10 with error correction by the composite method to construct chemographs

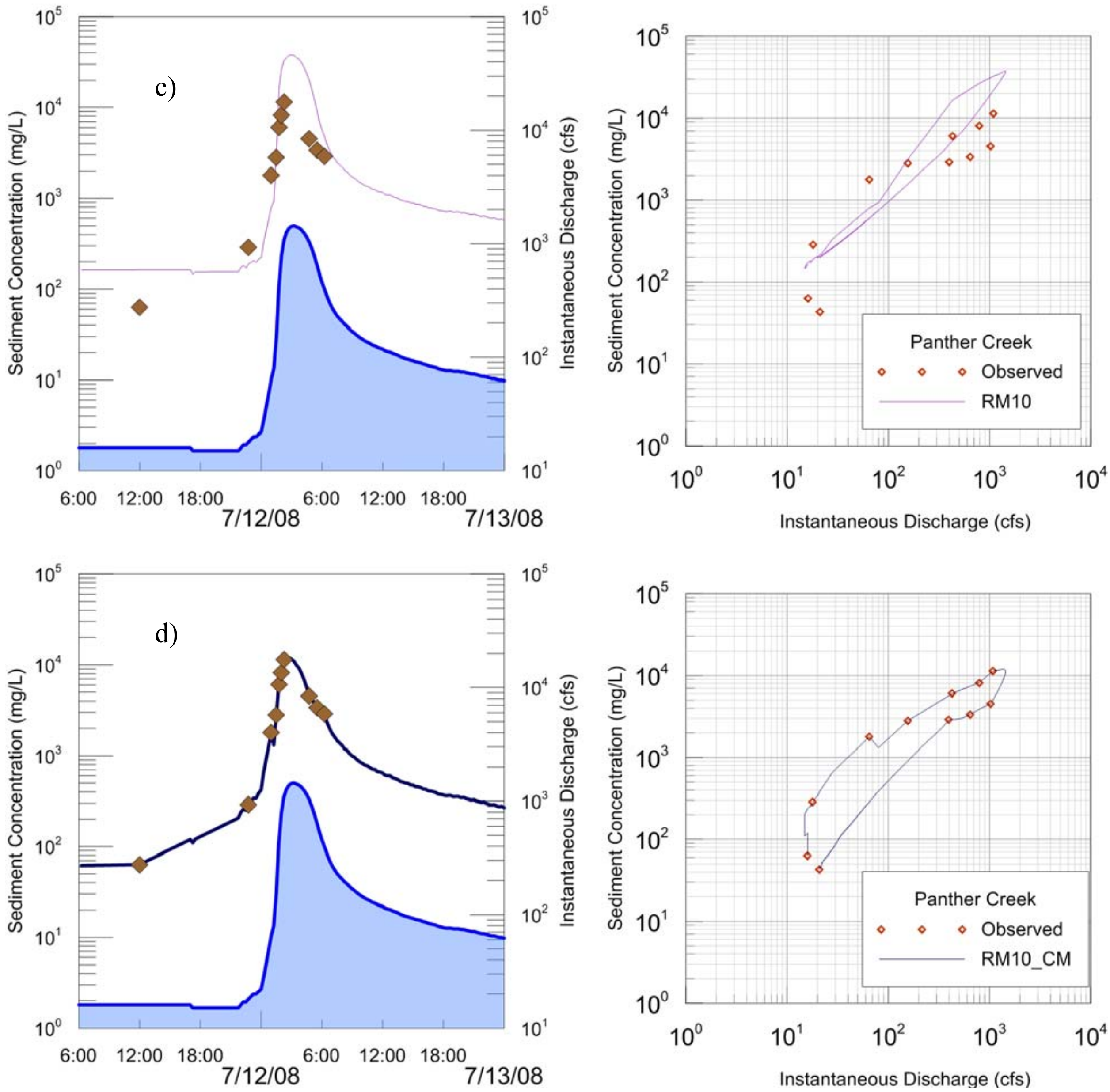

Figure 5-5. (concluded)

In order to quantify the magnitude of differences in predicted concentrations among the different methods of concentration estimation, concentration duration curves were created for all chemographs constructed for the entire study period. The simulated SSC duration curves for Cox Creek are presented in Figure 5-6. There is strong agreement in the predicted concentrations within the inter-quartile range (25th-75th percentile) but more variation in predicted concentrations at the extreme lowest and highest concentration percentiles.

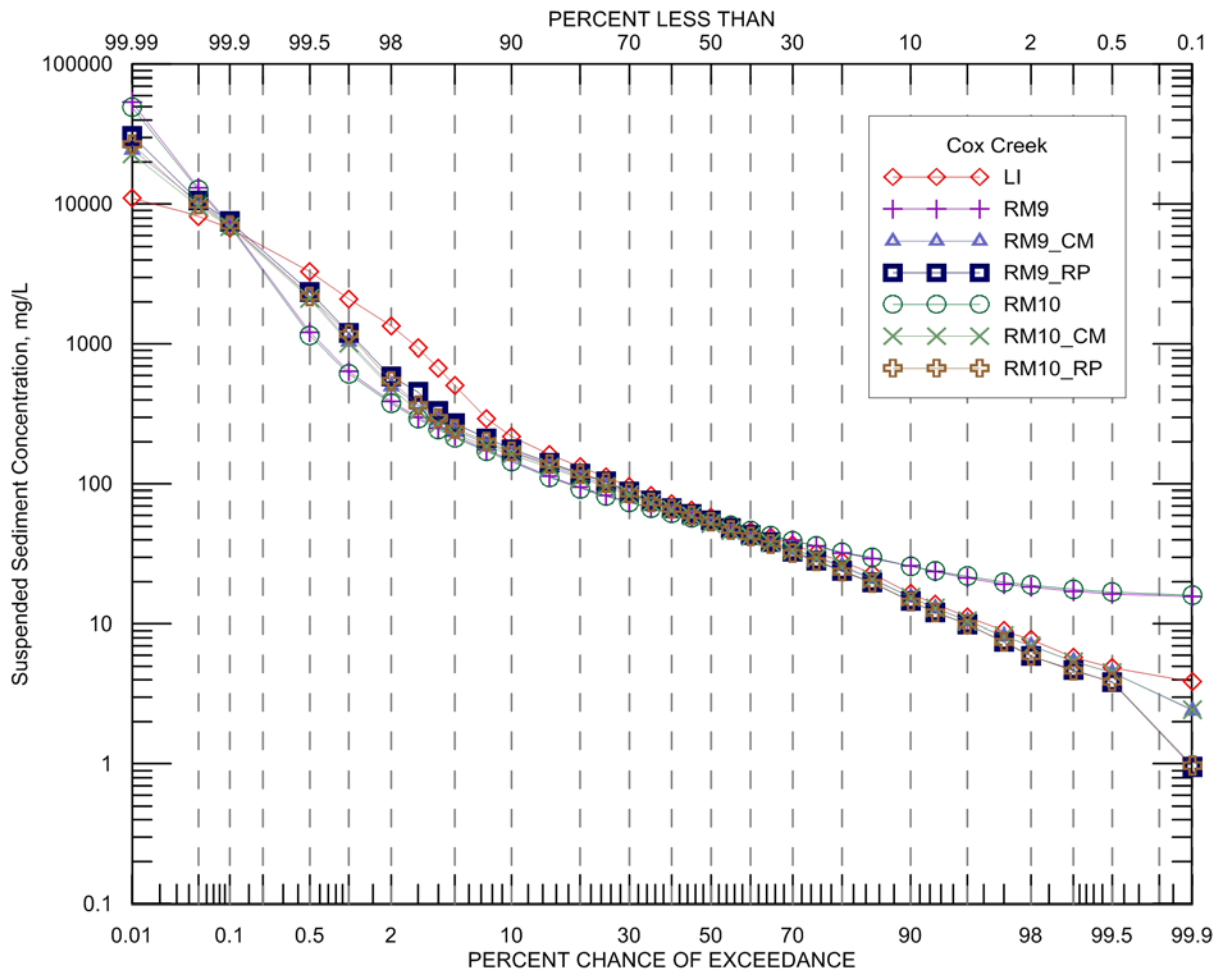

Figure 5-6. Suspended sediment concentration duration curves created from various chemograph construction methods, WY 2000-2009

Based on evaluation of SSC duration curves at all study sites, the concentration estimation techniques displayed the following characteristics:

- At the extremely low concentration percentiles, the two regression models predicted nearly identical concentrations which were generally higher than the estimates produced by error correction or linear interpolation.

- Error correction by the RP method predicted the lowest concentrations at the lowest percentiles for SSC.
- At all sites except Court Creek, linear interpolation and error corrected regression models (both CM and RP) predicted similar concentrations for the lowest 75-80% of all concentration percentiles. At Court Creek LI predicted higher concentrations until approximately the 50th percentile, and then predicted similar concentrations to error corrected regression models for the lowest 50% of all predicted concentrations.
- For the highest 10-25% of concentrations, LI generally predicts the highest concentrations, regression models predict the lowest concentrations, and the error corrected regression model estimates fall in between. This behavior changes at the extreme highest concentration percentiles (less than 1st percentile) where regression models predict the highest concentrations, LI the lowest, and error corrected models are again in between.
- At the highest concentration percentiles, RP error correction predicted higher concentrations than concentrations predicted by CM error correction.

Many of these behaviors make sense considering the nature of chemograph construction for the different methods. Because regression models and error correction techniques will extrapolate values, at the extreme highest percentiles their predictions were higher than those estimated from linear interpolation. Similar results are not seen at the lowest concentration percentiles, however, because low concentrations typically occur during low flows and are not subject to rapid changes of short duration, like experienced with high flows. Furthermore, the three smallest study sites experienced significant periods of no flow, at which time the regression models will predict concentrations based on seasonality and perhaps long-term time trends.

Selected concentration percentiles are presented in Figure 5-7 for Court Creek. Referring to the summary statistics for observed sediment concentrations at Court Creek (Table 3-8) underscores the bias storm sampling can introduce to summary statistics as more than 10% of observed samples had concentrations in excess of 2,176 mg/L. Yet these chemographs (Figure 5-7) suggest that sediment concentrations at or above this level occur less than 1% of the time. The impacts of storm sampling on summary statistics can also be seen in the median concentrations. The median of observed samples at Court Creek was 120 mg/L, but the median based on continuous estimates of concentration were less than 30 mg/L.

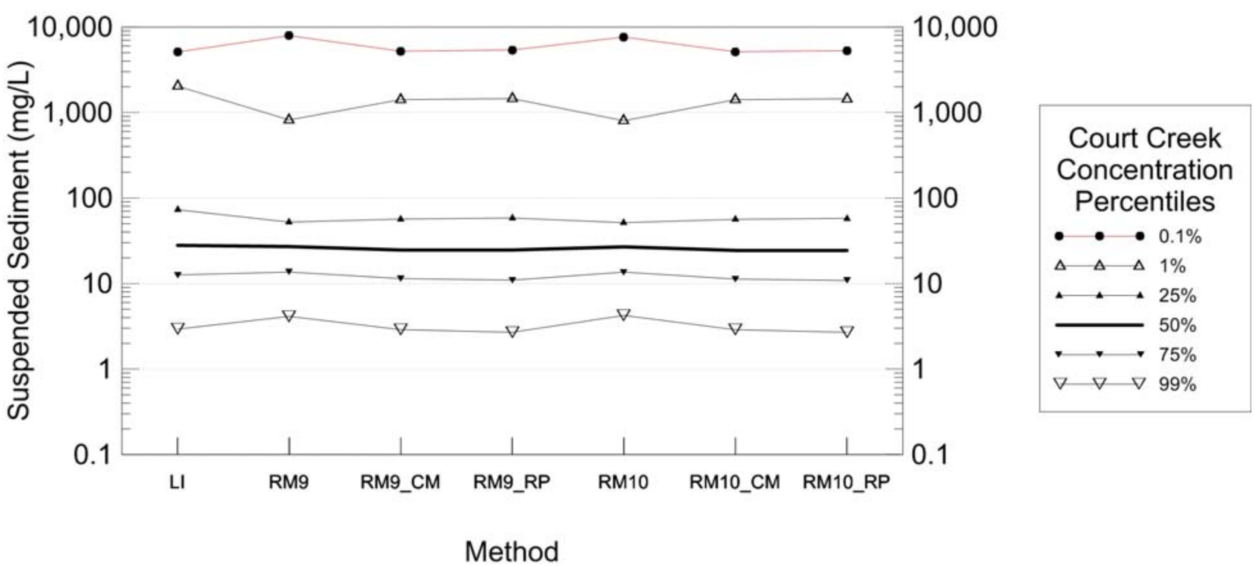

Figure 5-7. Selected percentiles of suspended sediment concentration as determined by various chemograph construction methods, WY 2000-2009

While concentration duration curves and plots of concentration percentiles depict the variation in predicted concentrations, they do not identify which chemograph most closely represents actual stream concentrations. Validation results are needed to assist in that assessment.

Validation data statistics for Court Creek and North Creek are summarized in Table 5-3. The shaded rows indicate the statistics for error corrected models. At both sites, the NSE and RSR stats were nearly identical for all error corrected models. The LI method essentially performed as well. While the regression models performed the worst. All methods were negatively biased, meaning constructed chemographs generally overestimated sediment at the time of sampling for the validation data set.

Table 5-3. Validation statistics for suspended sediment concentration estimates

	North Creek (Site #3)				Court Creek (Site #5)		
	NSE	Bias (%)	RSR		NSE	Bias (%)	RSR
LI	0.90	-2.46	0.31		0.92	-2.34	0.28
RM9	0.80	-5.88	0.45		0.80	-1.96	0.45
RM9_CM	0.91	-1.82	0.30		0.93	-1.10	0.26
RM9_RP	0.91	-1.15	0.30		0.93	-1.09	0.27
RM10	0.80	-5.51	0.45		0.80	-2.00	0.44
RM10_CM	0.91	-1.70	0.30		0.93	-1.11	0.26
RM10_RP	0.91	-1.03	0.30		0.93	-1.10	0.27

While the validation statistics for error corrected models are similar, one of the error corrections to RM10 was selected as the best representation because regression model statistics previously presented (Section 4.2.1) offer RM10 as a significant improvement over RM9. Even though the validation statistics for North Creek may suggest that concentrations estimated by the RP method are slightly less biased than CM error corrected estimates, this slight difference was not considered significant enough to be the deciding factor. To assist in the decision, the hydrographs and chemographs during collection of the validation data set were examined. The abrupt changes in concentration resulting from error correction by the RP method (Figure 5-4) was ultimately rejected as the best representation of stream concentrations because this technique can predict such drastic changes in SSC when there is no physical reason to expect these to occur. Therefore, the RM10_CM chemograph was selected as the best predictor of SSC and subsequently used to generate the 15-minute record of suspended sediment concentration necessary for TP Models 12 and 13.

5.2.2 Comparison of Phosphorus Concentration Estimates

Because TP samples were collected much less often than sediment, the poor performance of the LI method was even more magnified for the TP chemographs. Because TP is directly proportional to flow at these study sites (Figure 4-2), TP estimates were underestimated by the LI method when runoff events were not sampled and the previous and subsequent sample were both collected during lower flow conditions. Conversely, TP was often overestimated if the last sample collected during a runoff event still showed elevated concentrations but the stream returned to predominantly baseflow conditions and was not sampled again until the next routine monthly sample, which could be weeks later.

Due to the far fewer TP samples than SSC samples, the characteristics of chemographs constructed by the RP method were more evident in the TP predictions. At Panther Creek, RM10 predicted concentrations higher than the observed TP samples collected on both 7/9/2008 and 7/15/2008 (Figure 5-8). RM10 was error corrected using the composite method by applying linearly interpolated residual values (RM10_CM). Applying the residuals using the RP method can result in abrupt shifts in the chemograph at the mid-point between observations as can be seen during the receding limb on 7/12/2008 and during baseflow conditions on 7/18/2008. The magnitude of these abrupt shifts is determined by the difference in residual magnitude between adjacent observations.

Figure 5-8 also illustrates on 7/15/2008 how CM error correction can sometimes create a local minima in chemographs at observed samples during periods when flow and model residuals are both decreasing (i.e. simulated concentrations are more closely approaching observed). There is no physical reason to expect this sample to represent the absolute lowest TP concentration and then for TP concentrations to begin increasing immediately after sample collection. I think during these situations, RP error correction predicts more accurate concentrations because it allows sample concentrations to change gradually with flow as predicted by the regression equation but just shift the concentrations all up or all down. This can partially explain why RP error correction performed so well with the validations data set for SSC. Validation data was often collected within 24 hours of the primary data set, so serial correlation was still very strong and RP can at times better simulate concentrations closest to observed samples. During low flows, these differences between CM and RP methods of error correction can be important; however, these differences would have less of an impact during periods of frequent sampling and/or rapid changes in concentration. This may explain why Verma et al. (2012) found RP error correction performed so well with a dissolved constituent sampled daily, but the same performance is not seen in this study with particulate constituents that are irregularly sampled.

When visually comparing the various chemographs along with the accompanying streamflow record, it was noted that while RM10_CM can create local minima in its TP chemograph, RM12_CM and RM13_CM generally do not. This is a result of including in these models a suspended sediment chemograph as an explanatory variable which was constructed from a regression model corrected by nearly daily SSC observations. In general RM13 tended to predict lower peak concentrations during runoff events than RM12 which predicted lower peak TP concentrations than both RM9 and RM10.

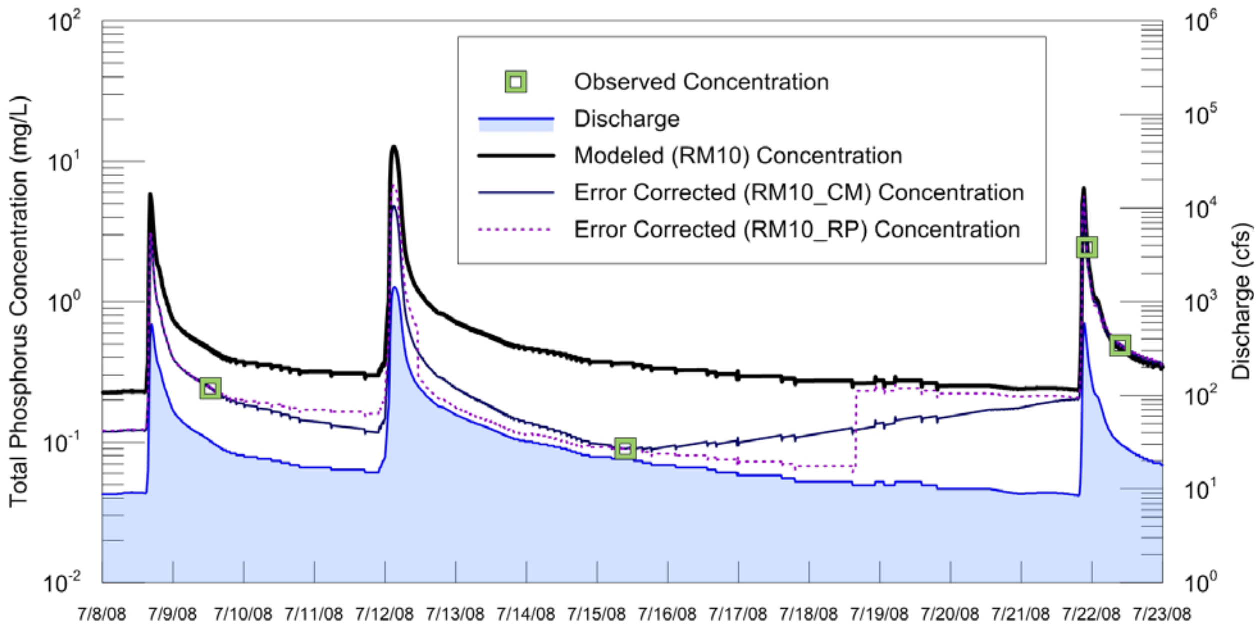

Figure 5-8. Comparison of TP chemographs at Panther Creek constructed using RM10 (thick black line), RM10_CM (thin solid line), and RM10_RP (thin dashed line)

The simulated TP duration curves for Haw Creek are presented in Figure 5-9. In general, there appears to be much greater variation in predicted TP concentrations than SSC but in actuality most of this variation is just a scale issue. At these study sites, TP concentrations typically span three orders of magnitude (0.01-10 mg/L) while suspended sediment concentrations vary over five orders of magnitude. The smallest study site, Cox Creek, exhibited the greatest variation in predicted TP concentrations.

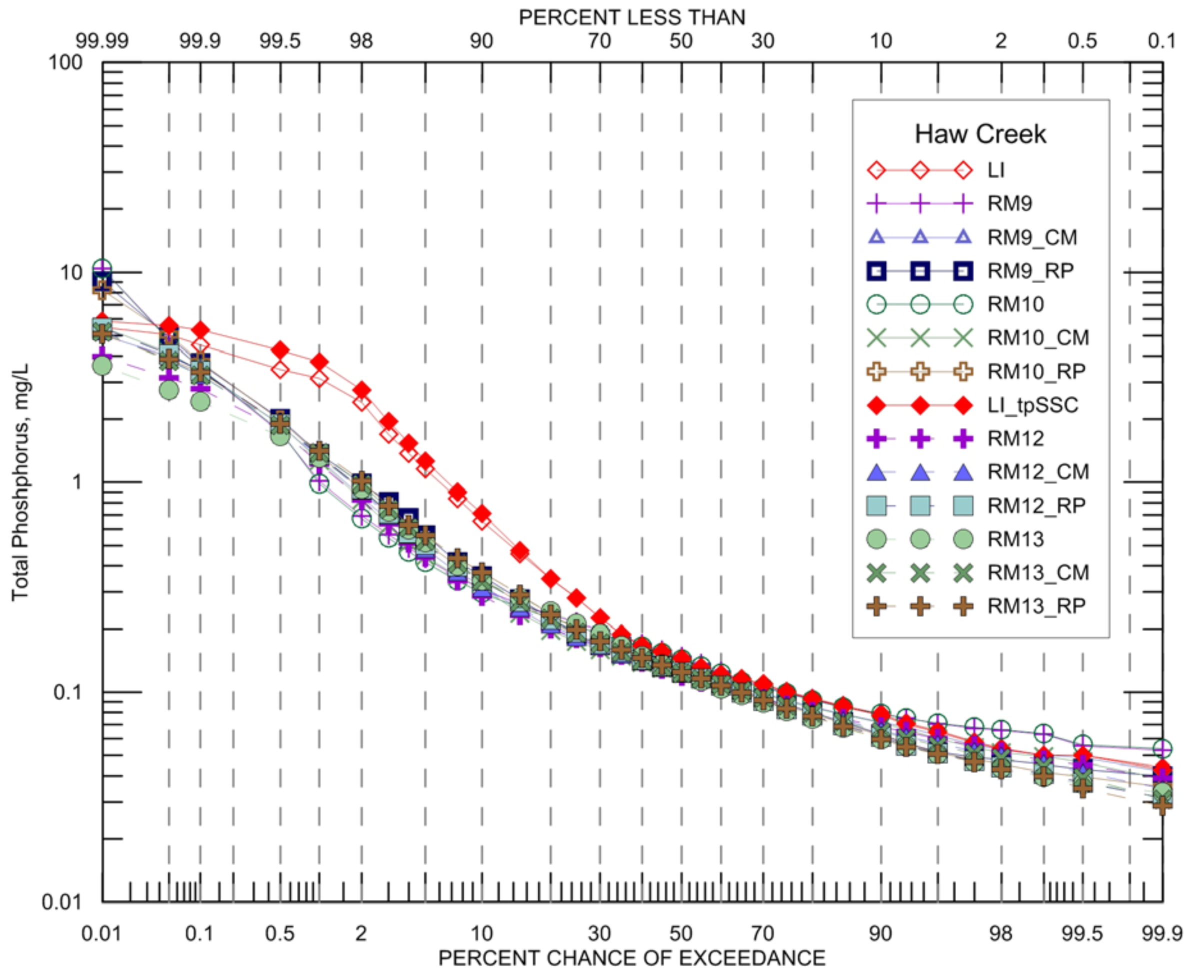

Figure 5-9. Total phosphorus concentration duration curves created from various chemograph construction methods, WY 2000-2009

Based on evaluation of TP concentration duration curves at all study sites, the concentration estimation techniques displayed the following characteristics:

- At the extremely low concentration percentiles, regression models 9 and 10 predicted nearly identical concentrations which were generally higher thanRM12 and RM13 TP predictions, which were in turn generally higher than the estimates produced by error correction or linear interpolation.
- Error correction by the RP method predicted the lowest concentrations at the lowest percentiles for TP.

- For the highest 30-50% of concentrations, LI methods predict the highest concentrations, regression models predict the lowest concentrations, and the error corrected regression model estimates fall in between. This behavior changes at the extreme highest concentration percentiles (less than 0.1 percentile) where the chemograph behavior is much more site and dataset specific. At all sites except North Creek, the highest TP predictions were made by RM9 or RM10; at North Creek RM9_RP predicted the highest TP concentrations.

Selected concentration percentiles are presented in Figure 5-10 for Panther Creek. Referring to the summary statistics for observed TP concentrations at this site (Table 3-9) the impact of storm samples on summary statistics is evident as more than 10% of observed samples had concentrations in excess of 3.30 mg/L. Yet these chemographs estimated TP concentrations at or above this level occur less than 0.5% of the time at Panther Creek. The impacts of storm sampling on summary statistics can also be seen in the median concentrations. The median TP concentration of observed samples at Panther Creek was 0.26 mg/L, but the median based on continuous estimates of concentration were less than 0.16 mg/L.

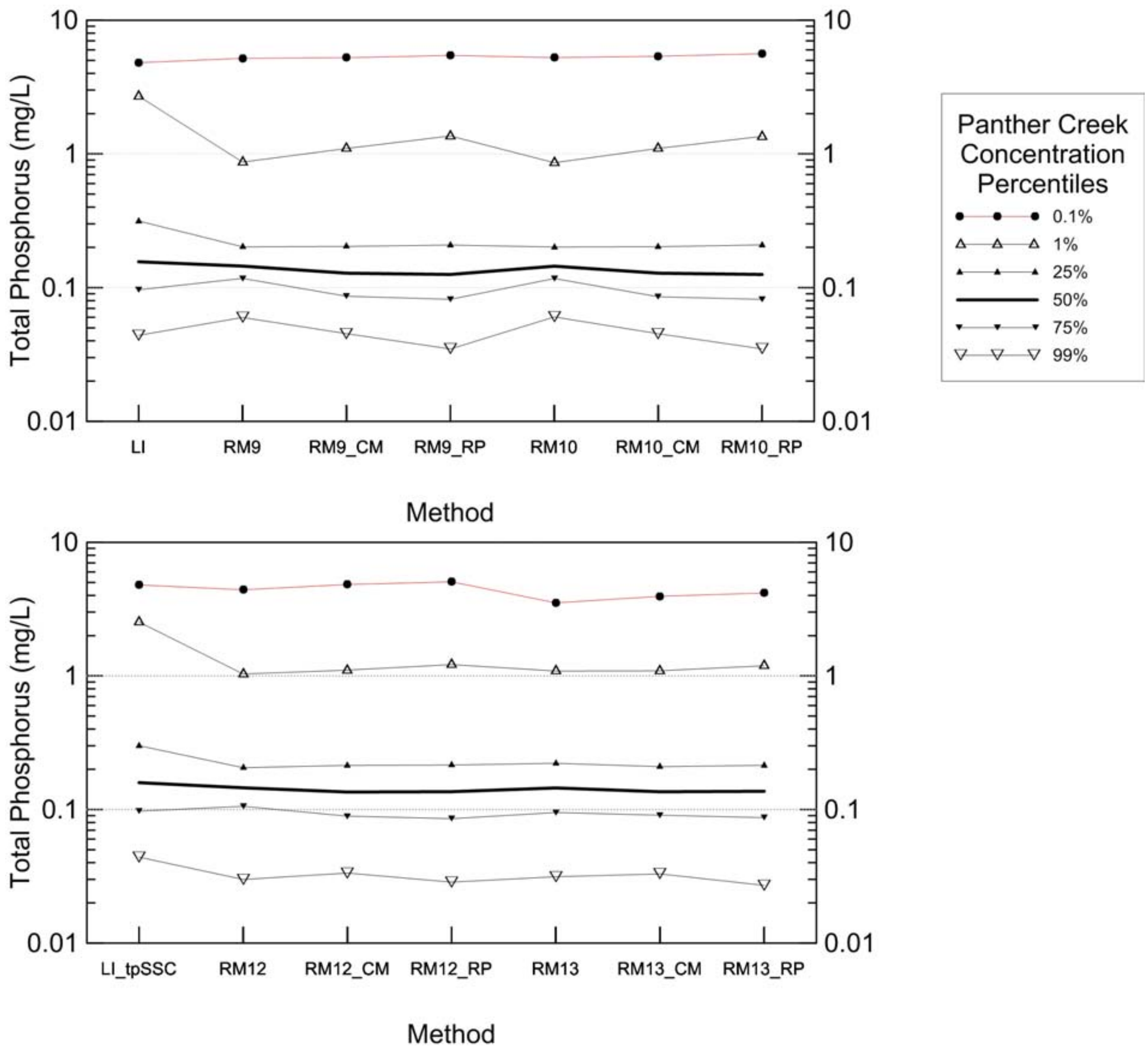

Figure 5-10. Selected percentiles of total phosphorus concentration as determined by various chemograph construction methods, WY 2000-2009

Validation data statistics for North Creek and Court Creek are presented in Table 5-4. The shaded rows indicate the statistics for error corrected models. The error corrected models performed better than their uncorrected regression model counterparts and RM12 and RM13 performed better than RM9 and RM10. At North Creek, the NSE and RSR stats indicated the error corrected models for RM12 and RM13 performed better than the error corrected models for

RM9 and RM10. At Court Creek there was essentially no difference in the performance of the various error correction approaches. At both sites the LI methods performed the poorest. It should be noted that LI_tpSSC is also linear interpolation between observed samples; it is simply using a smaller TP dataset. The fact that LI performed so well with SSC yet so poorly with TP is a reflection of the impact of the difference in the number of observed data points for the two constituents.

Table 5-4. Validation statistics for total phosphorus concentration estimates

	North Creek (Site #3)				Court Creek (Site #5)		
	NSE	Bias (%)	RSR		NSE	Bias (%)	RSR
LI	0.46	10.35	0.74		0.69	10.67	0.56
RM9	0.69	9.14	0.56		0.77	7.23	0.48
RM9_CM	0.77	2.50	0.48		0.93	3.16	0.27
RM9_RP	0.73	2.74	0.52		0.91	3.22	0.30
RM10	0.69	8.34	0.56		0.77	5.71	0.48
RM10_CM	0.77	2.81	0.48		0.93	2.27	0.27
RM10_RP	0.73	3.02	0.52		0.91	2.24	0.29
LI_tpSSC	0.34	9.86	0.81		0.62	11.87	0.62
RM12	0.80	6.46	0.45		0.90	6.94	0.31
RM12_CM	0.86	3.83	0.37		0.93	5.40	0.26
RM12_RP	0.84	3.60	0.40		0.93	5.04	0.26
RM13	0.79	-1.37	0.46		0.89	1.88	0.33
RM13_CM	0.86	4.45	0.37		0.92	5.91	0.28
RM13_RP	0.84	4.06	0.40		0.92	5.54	0.29

Based on the validation results, TP predictions were positively biased, meaning constructed chemographs tended to underestimate TP concentrations. The percent bias is greater for TP predictions than SSC predictions, but this is not a major concern due to the magnitude of TP concentrations. According to Table 5-2, half of all samples in the validation data set were less than or equal to 0.12 mg/L, so even a predicted concentration of 0.11 mg/L would result in a bias of 8.3%for that observation.. Therefore, the percent bias results are not as worrisome as they would be for flow or sediment predictions whose ranges are much larger.

Based on validation results and model performance statistics (Section 4.2.2), RM12_CM was selected as the best method for constructing TP chemographs. Regression model 12 requires a suspended sediment chemograph, and not all monitoring programs will have the benefit of such a robust SSC dataset. In those situations, RM10_CM would be the best chemograph technique to utilize when continuous records of suspended sediment concentration are not available. The similarity in annual TP loads for all five study sites estimated by these two methods is displayed in Figure 5-11, along with a line of perfect agreement.

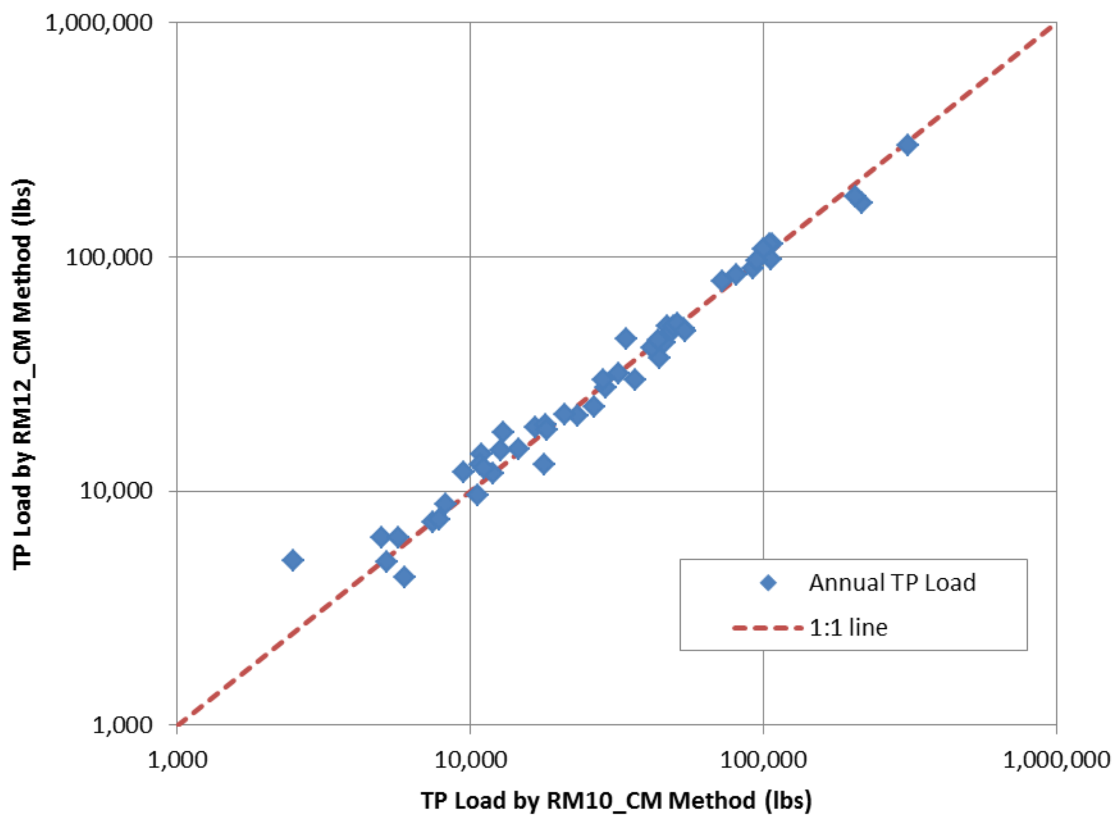

Figure 5-11. Comparison of annual TP loads computed when SSC is (y-axis) and is not (x-axis) used as a predictor variable. Dashed line represents line of perfect agreement.

These concentration duration curves and percentile graphs are an interesting way to present differences in the frequency of concentration predictions but they do not quantify the cumulative impact of a sequence of over- and/or underestimations of concentrations or the flow conditions under which they occur and the subsequent effect on load calculations.

5.2.3 Comparison of Sediment Load Estimates

Because RM10_CM was selected as the best representation of actual in-stream suspended sediment concentrations, the load calculated using this chemograph is considered the "true" load for subsequent comparisons. Box-and-whisker plots summarizing by method the annual deviation from the “true” load are presented in Figure 5-12. RM9 and RM10 produced the greatest deviation in annual loads for all study sites.

Sediment chemographs constructed by LI produced annual loads that ranged from 31-130% of the true load and were on average 93% of the true load at the two Sangamon sites and 106% of the true loads at the three Spoon watershed sites. RM9 and RM10 produced annual load estimates that averaged 156% and 148%, respectively of the true annual loads. The error corrected regression models RM9_CM (106%), RM9_RP (109%), RM10_RP (101%), all produced similar annual sediment loads.

When looking at load estimates for specific water years, RM9 and RM10 greatly overestimated sediment loads at all sites during WY 2002. RM9 loads were 143-582% of “true” load, while RM10 loads ranged from 140-540%. These regression models also overestimated loads during other high flow years at the study sites. During WY2008, RM9 and RM10 respectively averaged 367% and 324% of the true sediment loads at the two Sangamon sites. During WY2009, RM9 and RM10 loads averaged 360% and 339% respectively of the true sediment loads at the three Spoon sites. RM9 and RM10 also significantly underestimated sediment loads at all sites during years of below normal flow, such as 2003 (37-72% of true) and 2006 (36-96% of true).

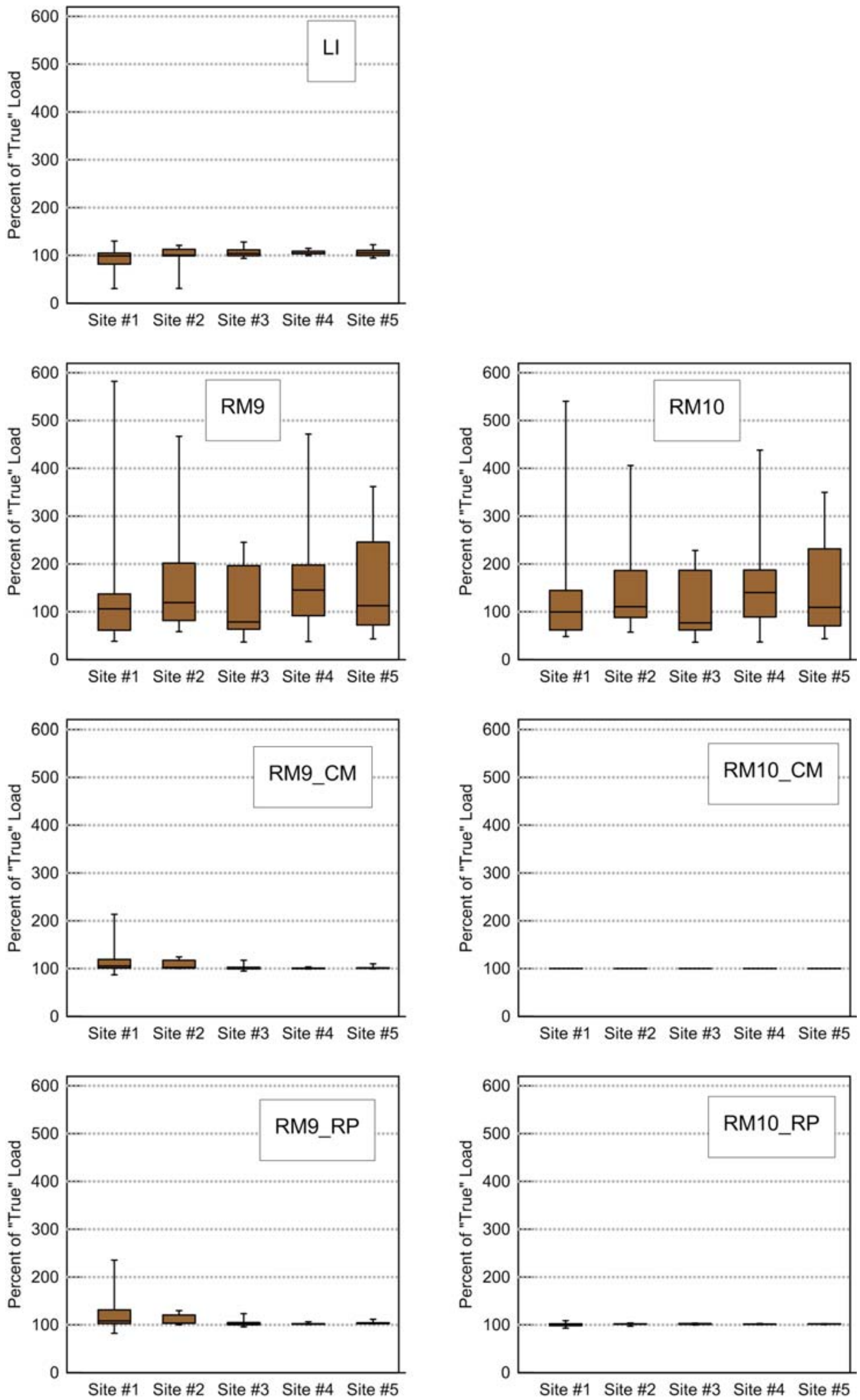

Figure 5-12. Annual suspended sediment load estimates by various chemograph construction methods as percent of "true" annual load (RM10_CM) for each study site

To explore the cumulative effect of these annual load estimate errors, Figure 5-13presents the total percent difference from the "true" load for the entire ten-year study period. The cumulative difference from true load was most pronounced at the 2 smallest sites, although for the two regression models the error was comparable to the error at the 2 largest Spoon sites. At the two smallest sites in the Sangamon watershed, LI produced suspended sediment load estimates that were approximately half of the true load, while regression model estimates were nearly triple the true estimate of sediment load for WY2000-2009. North Creek (site #3) appeared to be the least affected by choice of chemograph construction method as it showed the least deviation from true load over the study period. At North Creek RM9 and RM10 predicted SS loads that were 64% and 55% higher, respectively, than the true load. Error corrected and LI methods were all within 6% of true loads at the three Spoon watershed sites.

It is important to note that while concentration duration curves suggest RM9 and RM10 predict higher suspended sediment concentrations than the other methods at only the extremes, the impact these concentrations have on load estimates is striking and reinforces the overwhelming contributions large runoff events of extremely short duration can have to annual and ten-year load estimates.

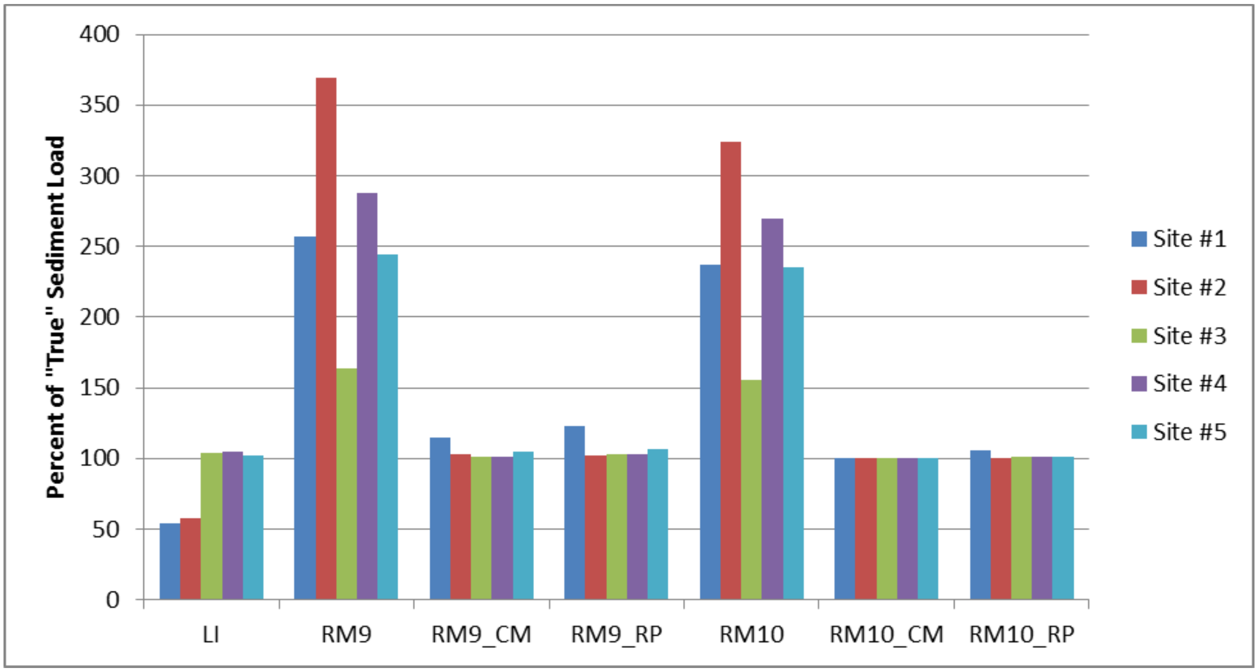

Figure 5-13. Ten-year (WY2000-2009) suspended sediment load estimates by various chemograph construction methods as percent of "true" load (RM10_CM) for each study site

5.2.4 Comparison of Phosphorus Load Estimates

With RM12_CM selected as the best representation of actual in-stream TP concentrations, the loads calculated using this chemograph are considered the "true" loads for subsequent comparisons. Box-and-whisker plots summarizing by method the annual deviation from the "true" load are presented in Figure 5-14. RM9 and RM10 again produced the greatest variation in annual loads at all five study sites.

Total phosphorus chemographs constructed by LI produced annual loads that ranged from 54-216% of the true load and were on average 9% higher than the true load at the five study sites. Regression models 9, 10, 12 and 13 produced annual load estimates that averaged 123, 120, 102, and 90% respectively of the true annual TP loads. The error corrected regression models all produced similar annual TP loads, averaging 95-105% of the true loads.

During the above normal flows of WY2002, regression models 9, 10, and 12 estimated TP loads that averaged 192, 186, and 105% respectively of the true annual loads while RM13 averaged annual TP loads that were only 85% of true. Similar to the suspended sediment models, RM9 and RM10 also significantly underestimated TP loads at all sites during years of below normal flow, such as 2003 (61% of true) and 2006 (77% of true).

The two smallest sites, Cox and Panther, showed the greatest annual deviation from true load. The underestimation of annual TP loads by the LI method was most pronounced at these two sites, and the deviation from true loads by the error corrected methods was greater at these two sites in the Sangamon watershed than at the three Spoon watershed sites.

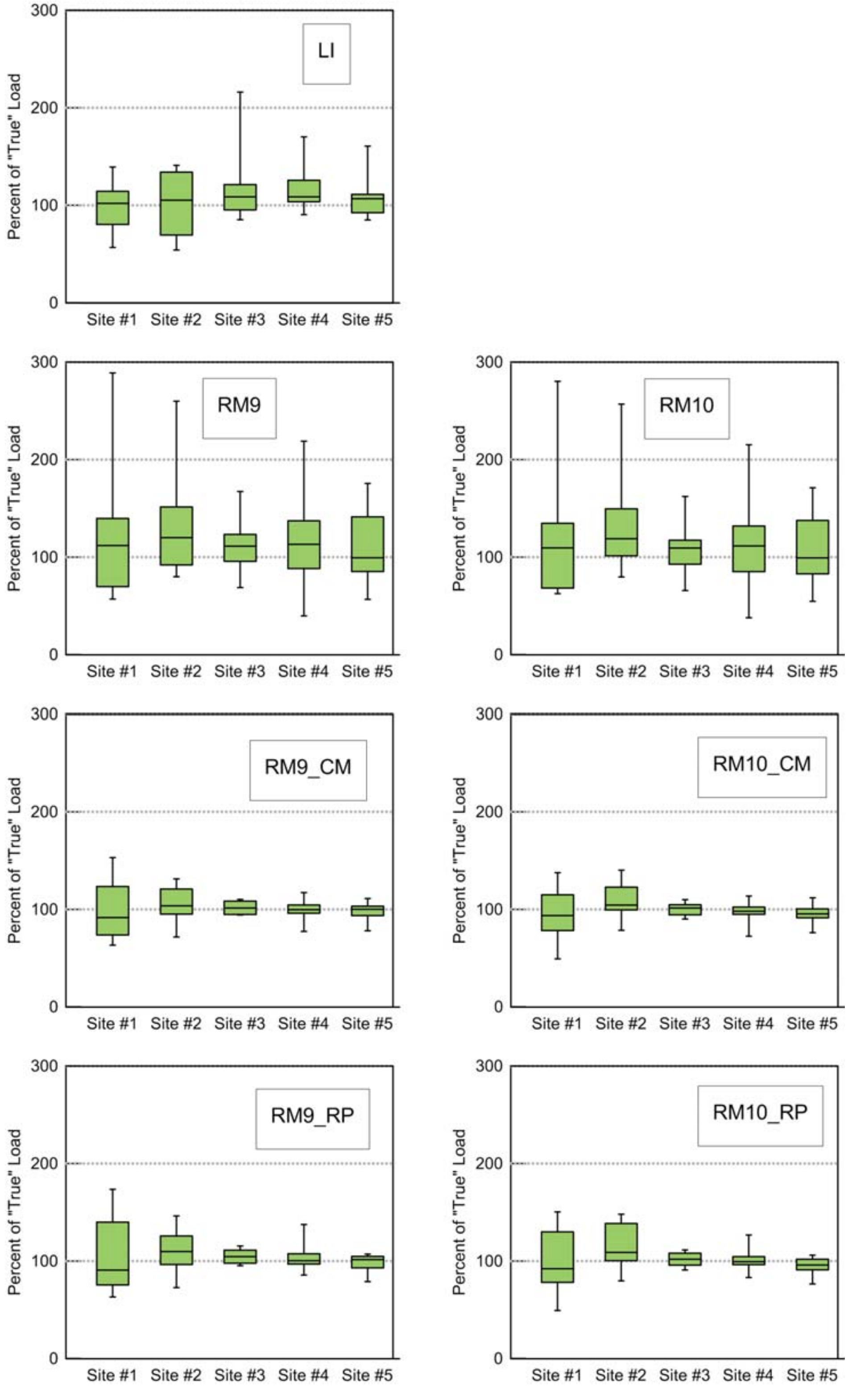

Figure 5-14. Annual total phosphorus load estimates by various chemograph construction methods as percent of "true" annual load (RM12_CM) for each study site

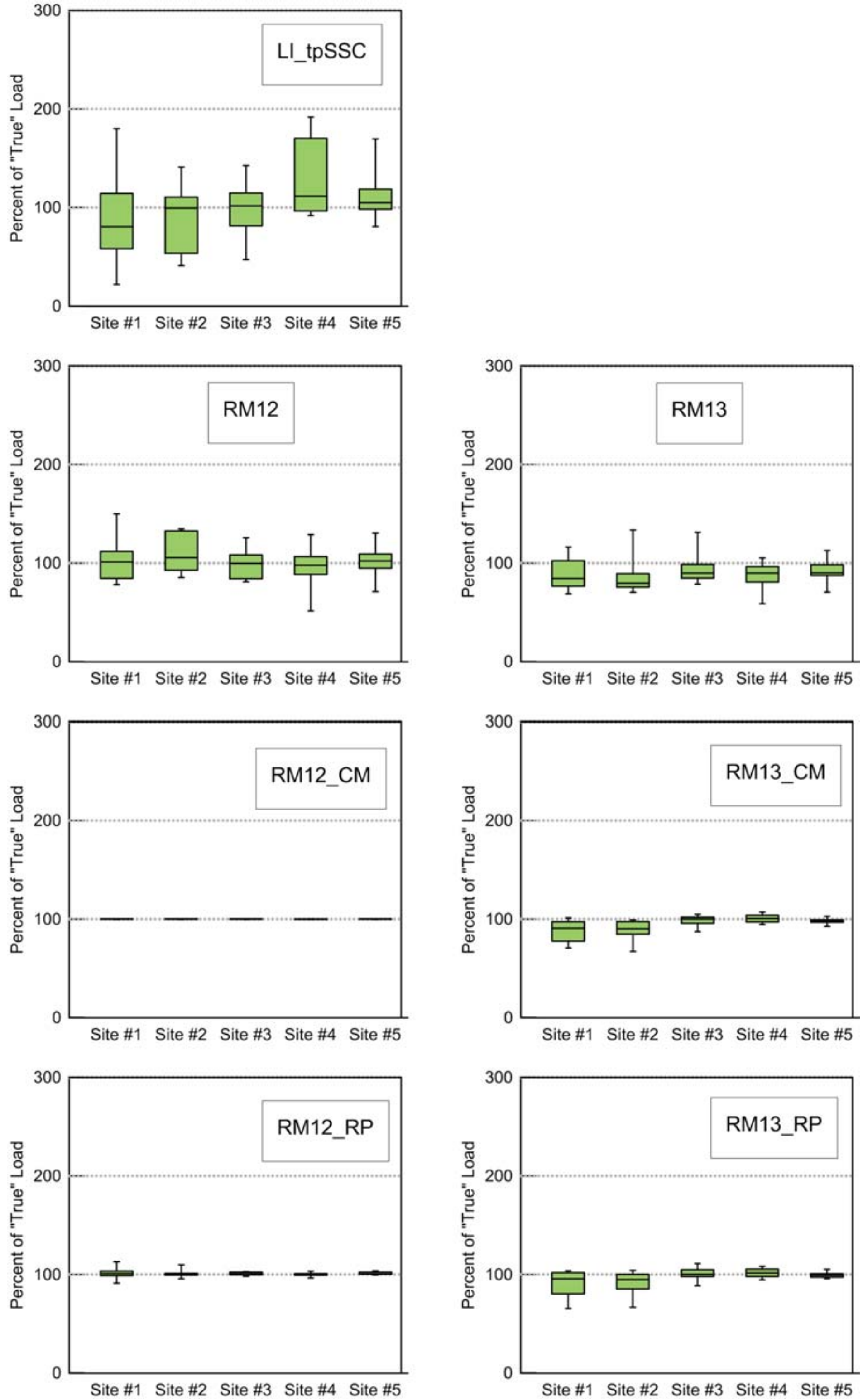

Figure 5-14. (concluded)

Figure 5-15 presents the total percent difference from the "true" TP load for the entire ten-year study period. The cumulative difference from true load was most pronounced at the two smallest sites in the Sangamon watershed, where LI produced TP load estimates that were approximately 20% lower than the true load. At the Sangamon watershed sites regression model 9, 10 and 12's load estimates were higher than the true estimate of TP load for WY2000-2009; RM13 predicted a lower TP load. At the Spoon sites, the LI method was 10% higher than true load. The LI_tpSSC chemograph was constructed using the same method as LI, but the observed dataset was just smaller. The load estimates using this smaller observed dataset to linearly interpolate concentrations resulted in TP loads 30% lower than true at Sangamon sites and 12% higher than true at Spoon sites. It is interesting to note the impact on ten-year load estimates at these small sites that results from reducing the number of observed samples by 11% (Table 4-1).

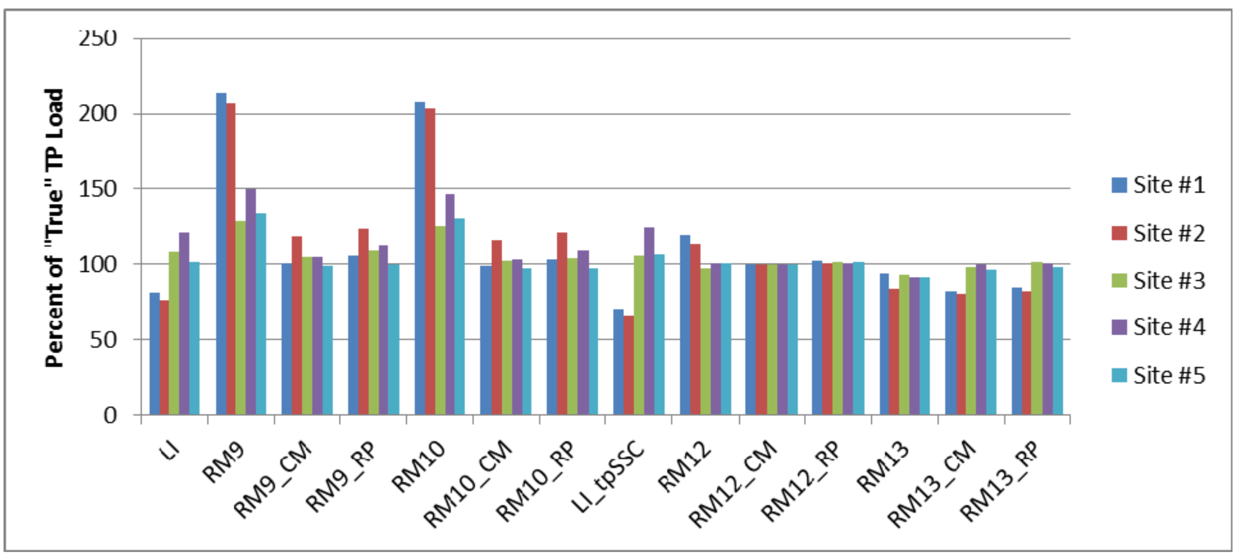

Figure 5-15. Ten-year (WY2000-2009) total phosphorus load estimates by various chemograph construction methods as percent of "true" load (RM12_CM) for each study site

5.3 Summary and Conclusions

This chapter quantified the differences between the four concentration estimation techniques and the effect on the subsequent load calculations. For both suspended sediment and total phosphorus, load calculations by error corrected regression models produced estimates that were the most precise and least biased of the four estimation techniques evaluated for these study sites. The method of error correction was not as critical as the act of error correction itself. Based on findings in this study, the following recommendations can be made:

1. Using regression models for load calculation without error correction is strongly discouraged as uncorrected regression models can create large overestimation of loads.
2. For suspended sediment load estimates, RM10_CM is a suitable alternative to the worked record approach as it can account for seasonality, hydrographic position, and streamflow without requiring subjective estimates of instantaneous measures of concentration. This approach is also objective and completely reproducible.
3. When possible, SSC should be included in regression models of TP concentration. Especially in studies such as this one where SS is collected more frequently and can be very helpful in predicting particulate concentrations.
4. For total phosphorus load estimates, the addition of SSC as a predictor variable improved regression model performance, while error correction of these models further increased estimate accuracy. RM12_CM is the recommended approach for TP load calculations when concurrent SSC is available for model building.
5. When concurrent SSC samples are not available, RM10_CM is the recommended approach for TP load estimation.

6. Even with targeted event sampling, linear interpolation is not recommended for creation of TP chemographs on small, flashy streams with significant runoff events of short duration. Linear interpolation proved to be a more suitable approach for sediment chemograph construction at the larger study sites due to the differences in stream response as a function of drainage area, as well as the extremely high number of observed samples,
7. Summary statistics derived from water quality samples whose collection included deliberate targeting of runoff events should not be interpreted as representative of the population of actual stream concentrations.

Another aspect to consider when interpreting the results of this chapter is the impact of the size of the datasets. For example, North Creek had the smallest errors in suspended sediment load estimates (Figure 5-13) and also had the highest number of SSC samples, followed by Haw, Panther, Court, and Cox. The difference in the number of samples at these sites was significant; North Creek had 40% more SSC samples than Cox Creek. TP samples showed a similar discrepancy. The Spoon watershed sites had approximately 33% more TP observations than the Sangamon sites. So, some of what we are seeing in the results is most likely due not only to drainage areas but differences in the number of samples available for model building and error correction.

6. SEDIMENT AND PHOSPHORUS LOADS

The previous chapter identified the best method for estimating loads at the study sites. Using these methods allows for further analyses to determine the proportion of sediment and phosphorus loading as a function of time and flow at the five study sites.

6.1 LOADING RESULTS

Despite its overall strong performance, the selected sediment chemograph (RM10_CM) at times predicted unrealistically high concentrations at both Cox Creek and Panther Creek. Upon closer investigation it was discovered that these were the result of a weakness in the regression model during certain hydrologic conditions. Both regression models 9 and 10 are heavily influenced by streamflow, and the predicted sediment chemograph closely mimics the shape of the hydrograph. Through the process of error correction, the numerous observed samples reshape the chemographs to better describe the sediment and phosphorus transport. When observed samples are missing during a runoff event, the simulated chemograph may not be reshaped in a realistic manner. One scenario where this can be problematic is during a runoff event comprised of multiple peaks, especially where the second event's peak flow is higher than the first. In this scenario, the regression models tended to predict higher peak concentrations during the second runoff event, but observed concentrations during these types of double-peak events do not follow that behavior. It is more common that concentrations during the first flush will be higher than concentrations during the second flood wave even if the second flood wave experiences higher peak flows. Typically when this scenario occurs, the observed samples adjust the concentrations down during the second peak. When no samples are collected during the second peak to adjust regression model estimates, the load estimates can be substantially over estimated.

This scenario occurred at Panther Creek on 9/14/2008 when its automatic pump sampler was full after sampling runoff events earlier in the week. Sampling resumed later that evening, but approximately 30 hours passed without the collection of a sediment sample during the largest runoff event of the year, Rather than utilize the unrealistically high SS concentrations and resultant loads in this chapter's subsequent analyses, I decided to error correct the regression model using a combination of the worked record approach and the composite method. Namely knowledge of the stream's behavior and professional judgment were used to estimate an instantaneous concentration which was treated like an observed concentration to correct neighboring estimates based on its regression model residual.

To estimate the single concentration, other storm events of similar magnitude were compared to determine a range of realistic concentrations. Secondly storm events containing double peaks were investigated to determine the typical concentration pattern during these types of runoff events. Overwhelmingly the peak sediment concentration during the second flood event was less than the peak concentration during the first runoff event even if the flood magnitude during the second event was higher.

For example at Panther Creek on 6/3/2008, a peak sediment concentration of 24,800 mg/L was observed during a runoff event, and less than 18 hours later a second flood wave passed with sediment concentration peaking at roughly 12,000 mg/L. This behavior was also observed during a double-peak runoff event in May 2009. Using this knowledge, a supplemental concentration was estimated during the second runoff event with a concentration roughly half of the peak concentration estimated during the first flood wave. This supplemental point was treated as an observed concentration and a regression model residual computed at its time of collection. This residual was used along with the others to error correct RM10's concentration estimates. The same approach was followed for a runoff event at Cox Creek in July 2008 where again the runoff event was severely under-sampled, and the only sample was collected on the rising limb and had a very large residual that only further increased subsequent regression model estimates.

A summary of the impact of these single events at Panther and Cox and their one-day load totals are presented in Table 6-1.TP loads estimated using RM12_CM chemographs were recomputed using the adjusted sediment chemograph as the continuous SS record in regression model 12. The resulting impact on TP loads was less pronounced than sediment but still dramatic. The incredible contribution of these individual under-sampled runoff events to load estimates at all time-scales are presented in Table 6-2.

Table 6-1. Comparison of predicted suspended sediment concentrations and loads following modified error correction

Site	Date	Original Predicted Concentration (RM10_CM)	Modified Prediction	Daily Load (RM10_CM)	Daily Load (modified error correction)
		(mg/L)		(tons/ac)	
1	7/12/2008	118,000	45,600	72,245	28,190
2	9/14/2008	63,100	9,700	149,730	44,290

Table 6-2. Percent reductions in sediment and phosphorus loads due to modified error correction

Site	Date	Daily	Monthly	Annual	Ten-Year
Suspended Sediment					
1	7/12/2008	61%	58%	28%	18%
2	9/14/2008	70%	70%	50%	31%
Total Phosphorus					
1	7/12/2008	44%	38%	11%	5%
2	9/14/2008	51%	49%	25%	13%

These modified load estimates are presented as a cautionary tale to stress the importance of reviewing and verifying any chemograph estimation method that extrapolates observed values. These scenarios at Panther Creek and Cox Creek also highlight the need for these error corrected regression models to be used with caution as they may not be suitable for load calculations on streams without sufficient event sampling; in fact, even with robust monitoring the role of professional judgment and knowledge of the streams is needed.

6.1.1 Annual Variation

The annual flow variation at the five study sites was presented in Figure 3-5. The year with the highest runoff at the two sites in the Sangamon watershed (2008) contributed more flow than the driest four years combined. The annual variations in SS and TP loads are presented in Table 6-3 and Table 6-4, respectively.

Table 6-3. Annual suspended sediment loads during study period, tons

	Site #1	Site #2	Site #3	Site #4	Site #5
WY2000*	8,560	4,414	6,735	21,382	26,010
WY2001	9,686	9,816	18,469	50,363	45,864
WY2002	25,876	43,953	27,991	44,407	63,942
WY2003	2,367	2,694	11,720	5,394	21,238
WY2004	4,672	7,780	2,082	10,910	7,132
WY2005	8,470	13,116	6,164	17,318	18,955
WY2006	3,066	2,342	3,795	5,673	7,831
WY2007	9,841	13,087	16,587	19,147	48,204
WY2008	112,927	106,479	20,926	16,878	41,407
WY2009	16,295	32,427	62,228	105,025	189,129

*Partial record

Table 6-4. Annual total phosphorus loads during study period, lbs

	Site #1	Site #2	Site #3	Site #4	Site #5
WY2000*	12,096	7,386	9,626	40,859	48,237
WY2001	14,455	12,959	27,665	83,526	78,982
WY2002	36,817	49,652	50,431	89,624	114,113
WY2003	5,062	6,317	19,099	17,864	44,780
WY2004	8,816	11,836	6,297	29,896	18,184
WY2005	13,016	20,947	15,170	43,140	44,012
WY2006	4,956	4,270	7,523	12,512	14,936
WY2007	18,834	21,162	31,894	50,630	108,164
WY2008	101,183	128,211	46,818	51,896	96,508
WY2009	22,969	29,690	97,941	181,437	301,538

*Partial record

While the wettest year contributed approximately 25% of the ten-year flow total at five study sites, the highest annual load contributed roughly 50% of the total ten-year sediment load and 44% of the total ten-year phosphorus load at the Sangamon sites (WY 2008) and 37 and 32% of the total sediment and phosphorus loads, respectively at the Spoon sites (WY 2009). At Cox Creek WY 2008 contributed more sediment than all other nine years combined. At all three Spoon sites the 2009 sediment load was greater than the load transported during the preceding six years combined. Similarly, WY 2008'sTP load at sites #1 and #2 was greater than the preceding seven years combined, and the Spoon sites TP load in WY 2009 was greater than the TP load transported from Water Years 2003-2007.

While the highest load producing years occurred during the years with greatest runoff, the proportion of loads generated per unit of runoff was not consistent during the ten-year study period, as illustrated for Panther Creek in Figure 6-1. For example, annual flow at Panther Creek in 2007 was only slightly higher than 2001's annual flow, yet SS and TP loads were approximately 50% higher in 2007 than 2001. Another example can be found in 2005 when flows were greater than in 2002, yet the SS and TP loads in 2005 were only 30-40% of 2002 loads at Panther Creek.

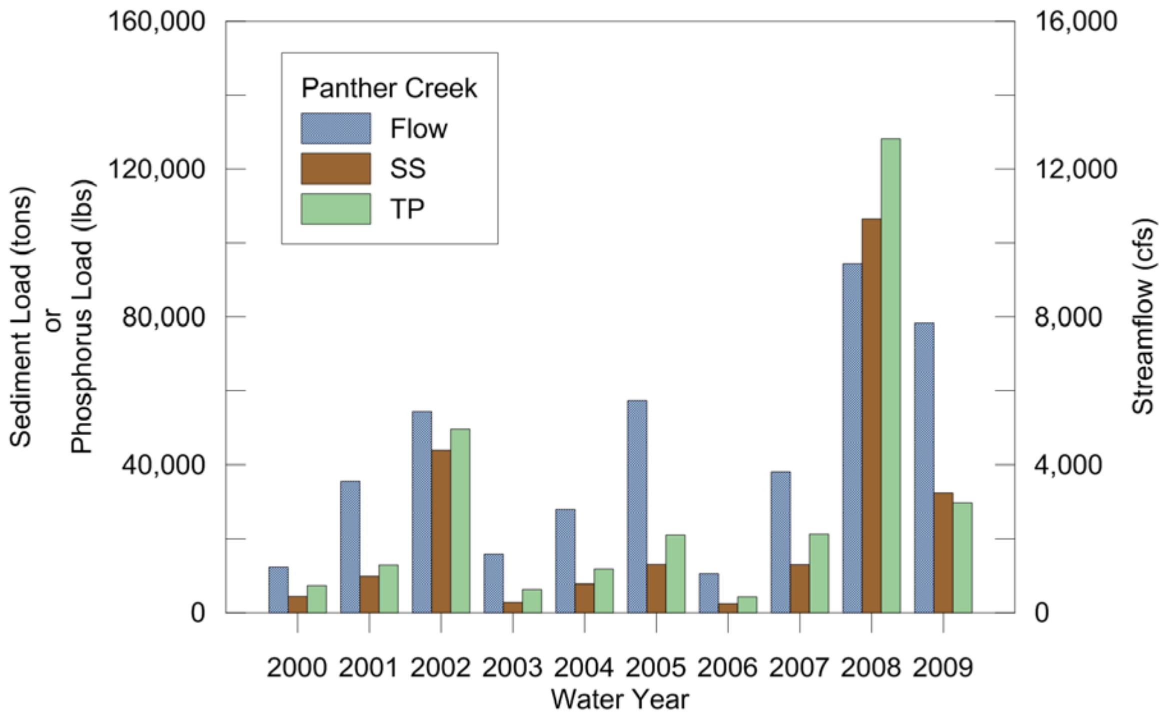

Figure 6-1. Annual sediment, phosphorus, and streamflow at Panther Creek

To remove the effect of watershed size in order to compare the sediment and phosphorus transported in each watershed, the loads were divided by each site's drainage area to compute yield or unit-area load. Annual median yields are presented in Table 6-5. Between the two Sangamon watersheds, Cox Creek produced higher sediment and phosphorus loads per unit area than Panther Creek. The Panther Creek watershed is approximately 33% larger than Cox Creek's watershed and while yields do tend to be inversely proportional to drainage area, this difference in yields may also be due to the higher percentage of agricultural land use in the Cox Creek watershed (93%) than the Panther Creek watershed (75%). Of the three Spoon sites, Haw Creek produced the lowest SS and TP yields. Based on annual medians, North Creek yielded more sediment per acre than the rest of the Court Creek watershed and produced lower phosphorus than the rest of the watershed; however, the median values do not reflect the annual variation. The gaged portion of North Creek composes 39% of the gaged portion of the Court

Creek watershed, yet annual loads from North Creek accounted for 26 to 55% of Court Creek's annual sediment and phosphorus loads. This variation in North's contribution to Court's annual loads is greater than the variation seen in North Creek's annual flows which accounted for 33 to 45% of Court's flow. North Creek's cumulative contributions over the ten-year study period were more proportional to their drainage area ratio; North's sediment and phosphorus loads were 38 and 36%, respectively of Court Creek's total loads and its flow was 40% of Court's ten-year total.

Table 6-5. Annual median yields, 2000-2009

Site ID	Site Name	Major Watershed	Suspended Sediment (tons/ac)	Total Phosphorus (lbs/ac)
1	Cox Creek	Sangamon	1.22	1.83
2	Panther Creek	Sangamon	1.08	1.61
3	North Creek	Spoon	0.83	1.37
4	Haw Creek	Spoon	0.52	1.32
5	Court Creek	Spoon	0.78	1.47

6.1.2 Seasonality

At the Sangamon watershed sites, the highest sediment loads occurred in late spring and early summer (Figure 6-2), although the median monthly sediment load in January and February was noticeably higher than the median monthly sediment loads for the preceding and following months. These loads are the result of runoff events during the winters of 2001, 2005, 2007 and 2008. These January and February high flow events are apparent in the bimodal peaks in maximum monthly streamflow for Panther Creek (Figure 3-6). Monthly total phosphorus loads at these two sites exhibited variation similar to that of the monthly sediment loads. Monthly total phosphorus loads at the Spoon watershed sites were typically highest in May (Figure 6-3), and the sediment loads at these three sites exhibited variation similar to the monthly TP loads.

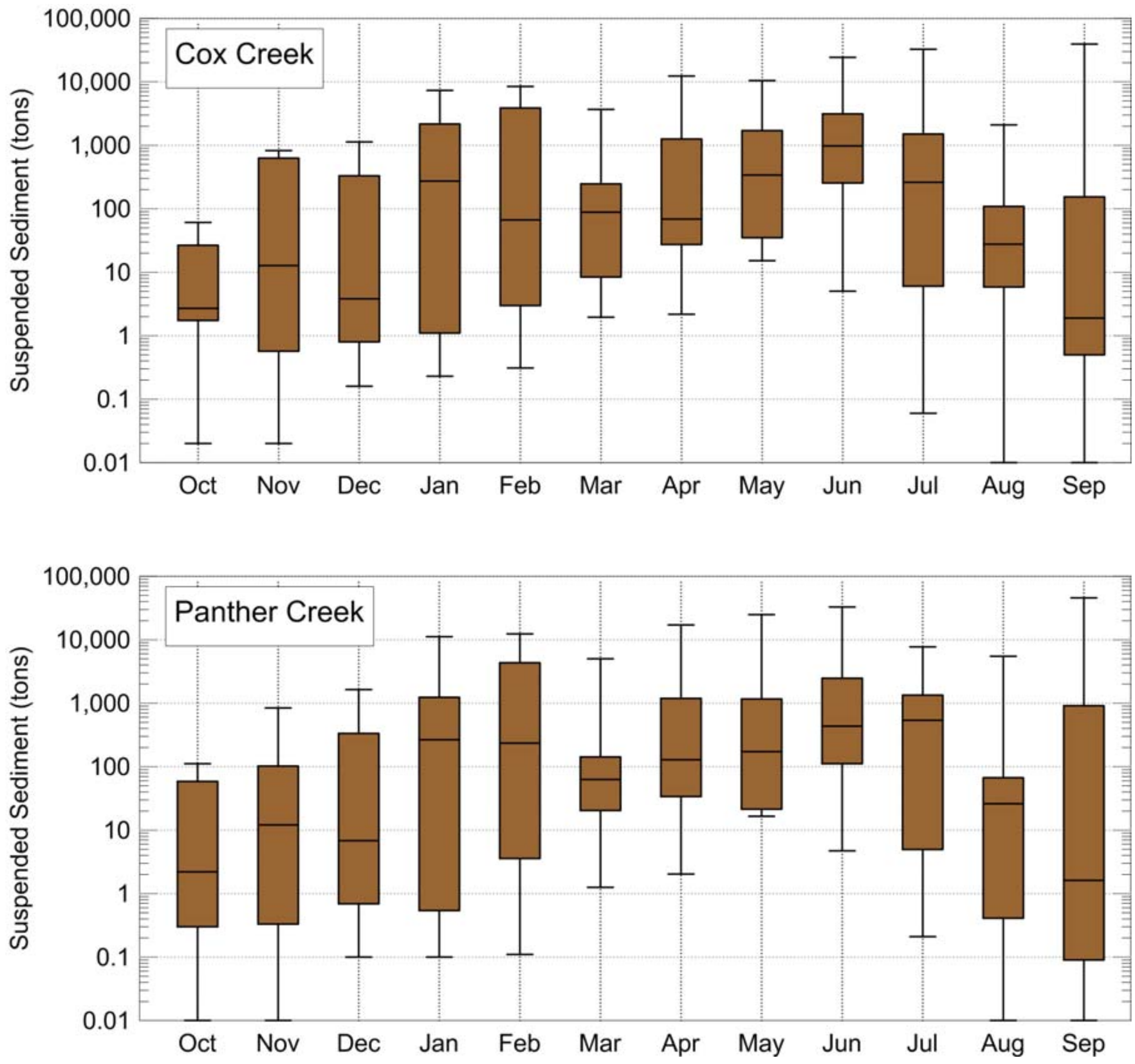

Figure 6-2. Monthly sediment loads at Sangamon watershed sites, Water Year 2000-2009

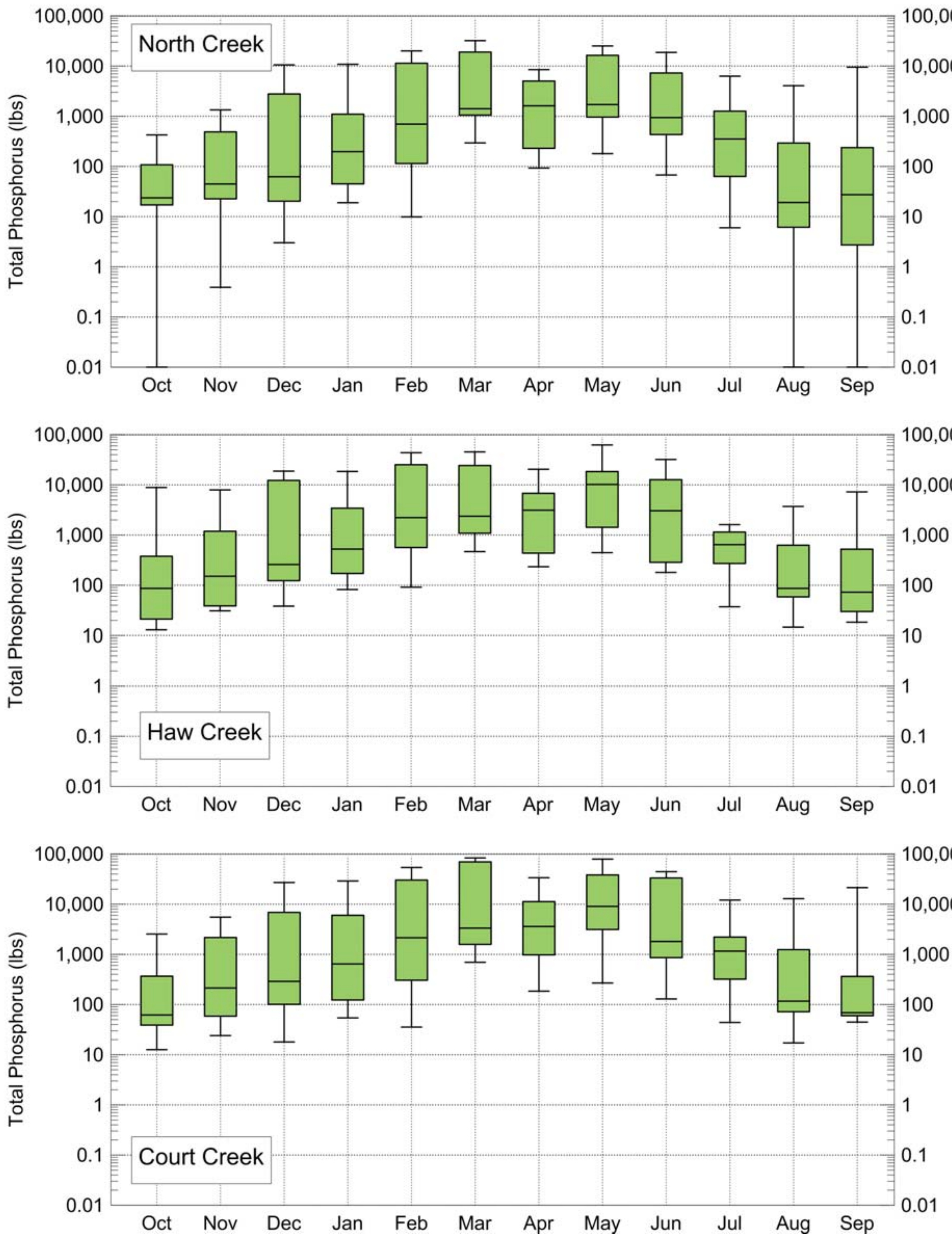

Figure 6-3. Monthly phosphorus loads at Spoon watershed sites, Water Year 2000-2009

At Panther and Cox, more than 25% of the total flow during the ten-year study period occurred in the months of May and June. Nearly 20% of the total phosphorus transport at these two sites occurred in the month of June; however, more sediment load was transported during the month of September than any other month of the year. More than 20% of the ten-year sediment load at these Sangamon watershed sites occurred in September, almost entirely due to the September 2008 runoff events. At the Spoon watershed sites, the month of May alone accounted for nearly 20% of the total flow and 25% of the total sediment and phosphorus loads at these sites.

To explore the seasonal distribution of flow and loads, the monthly loads were aggregated into seasonal totals for Winter (December, January and February), Spring (March, April, and May), Summer (June, July, and August), and Fall (September, October, and November). The effect of the September 2008 runoff events can be seen in the Fall sediment and total phosphorus loads being disproportionately higher than the Fall streamflows at the Sangamon sites (Table 6-6). Seasonal loads in the Spoon watershed sites (Table 6-7) were more proportional to flow contributions than at the Sangamon watershed sites.

Table 6-6. Seasonal distribution of streamflow and loads at Sangamon watersheds, percent

	Flow	Total Phosphorus	Suspended Sediment
Winter	29	22	20
Spring	37	26	24
Summer	20	33	34
Fall	14	19	22

Table 6-7. Seasonal distribution of streamflow and loads at Spoon watersheds, percent

	Flow	Total Phosphorus	Suspended Sediment
Winter	26	25	24
Spring	49	51	53
Summer	19	19	21
Fall	6	5	2

6.1.3 Duration Analysis

Mean daily streamflow, sediment loads, and phosphorus loads were sorted in descending order and the cumulative values plotted as a proportion of time for the ten-year study period (Figure 6-4). At the smallest study site, Cox Creek, approximately 80% of the streamflow occurs in 20% of the time, while greater than 90% of the sediment and phosphorus loads are transported in this small percentage of the study period. These figures dramatically illustrate the contribution of large runoff events of short duration to the total loads at these small study sites. The percentages of flow, sediment, and phosphorus transported during the study period are presented in Table 6-8, Table 6-9, and Table 6-10, respectively. During the ten-year study period, 5% of the time (the equivalent of roughly 183 days) accounts for approximately 50% of the flow, 91% of the total phosphorus load and more than 96% of the sediment load.

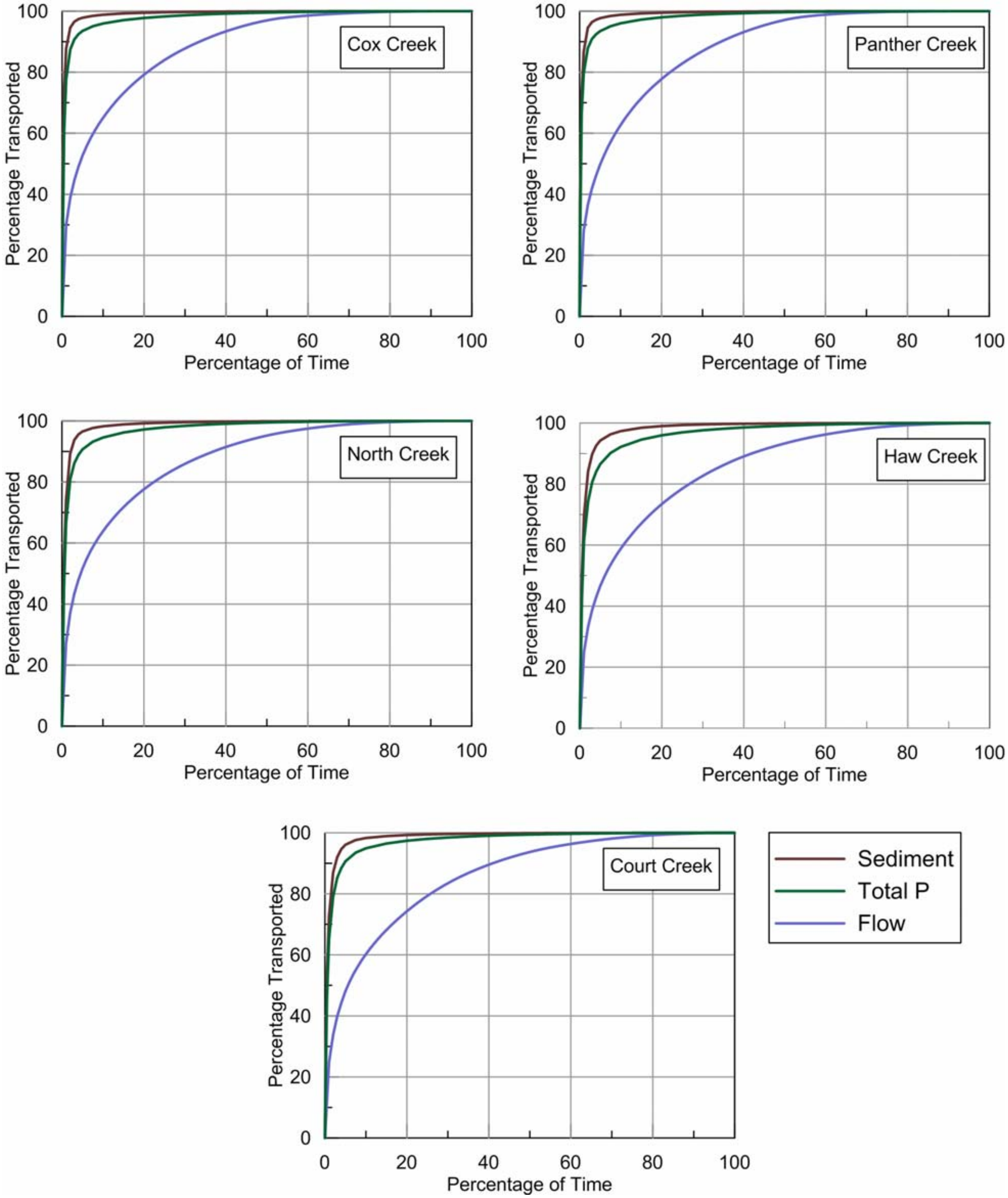

Figure 6-4. Percentage of flow, sediment, and total phosphorus transported as percentage of time, Water Year 2000-2009

Table 6-8. Percentage of flow transported during study period, Water Year 2000-2009

Percent of Time	Site #1	Site #2	Site #3	Site #4	Site #5
5	52.8	49.7	51.6	46.7	48.1
10	65.1	62.7	63.8	58.9	60.2
25	84.0	82.8	82.2	78.5	79.4
50	97.0	97.1	95.2	93.3	93.7

Table 6-9. Percentage of sediment transported during study period, Water Year 2000-2009

Percent of Time	Site #1	Site #2	Site #3	Site #4	Site #5
5	97.9	97.7	96.6	94.3	96.1
10	98.8	98.8	98.3	97.4	98.3
25	99.6	99.7	99.5	99.3	99.5
50	99.9	99.95	99.9	99.9	99.9

Table 6-10. Percentage of total phosphorus transported during study period, Water Year 2000-2009

Percent of Time	Site #1	Site #2	Site #3	Site #4	Site #5
5	93.5	93.5	90.7	86.6	90.7
10	96.0	96.0	94.6	92.2	94.9
25	98.3	98.5	97.9	96.9	98.0
50	99.6	99.7	99.5	99.1	99.4

The 183 days that comprised the top 5% of flows and loads were evaluated to determine whether these 183 days tended to represent only the wettest years of the study and whether the years of below normal flow were excluded from this top 5%. Overall the high flow years (2008 and 2009) did compose a disproportionate number of daily flows and loads, but every water year contributed to these 183 high flow days, and even the years of below normal flow contributed a few days that were within the highest 5% of flows and loads during the study period. This led to further investigation of the role of daily maximums in annual and total loads.

A single day accounts for 0.27% of the time each year. On average, at the Spoon watershed sites, the 1-day (daily) maximum accounted for 10% of the annual streamflow, 29% of the annual TP load, and 33% of the annual sediment load. The impact of a single day each water year during the study period is even more dramatic at the smaller Sangamon watershed sites where the daily maximums accounted for 11% of the annual streamflow, 36% of the annual TP load, and 43% of the annual sediment load.

It is interesting to note that the days with the daily maximum flows are not necessarily the same days which produce the daily maximum loads. For example at Cox Creek, daily maximum loads did not occur on the same date as the daily maximum flow in 2 out of the 10 years for TP loads and 3 out of the 10 years for sediment loads. Nearly half of the daily maximum sediment and phosphorus loads at North Creek did not coincide with the daily maximum flows each year. However, in 2003 and 2005 the daily maximum loads actually occurred one day prior to the date of daily maximum flow, which is not surprising given the tendency for the sediment and phosphorus chemographs to precede the hydrograph. At Court Creek the greatest 1-day sediment load occurred on 3/8/2009. The contribution of a single day during this runoff event produced 16% of the annual load during a very wet year, and accounted for more than 6% of the total 10-year sediment load. It should be noted that 3/8/2009 was not the greatest 1-day flow that year at that site. A greater 1-day flow occurred more than 2 months later on 5/15/2009.

6.1.4 Flow Normalization

As the previous section illustrated, a small percentage of time produces the majority of flow and an even larger proportion of loads. Because loads are a combination of flow and concentration, it is important to examine the relative contribution of flows to load totals because a decrease in loads may be more attributed to flow conditions (i.e., drought) than improvements in water quality. This concept of flow normalization is the subject of several large studies in the Chesapeake Bay watershed (Hirsch et al., 2010) and Mississippi River basin (Sprague et al,, 2011). The methods and techniques introduced in these studies are excellent tools for tackling this complicated issue but are unfortunately designed for even larger datasets (greater than 20 years in length with more than 200 samples collected per year) and are not intended for small, flashy streams. In this study, an attempt to account for this issue is made by normalizing loads by their flow contributions.

Dividing the load for a given time period by the streamflow during that period produces a flow-weighted concentration (FWC). Daily, monthly and annual FWCs were computed for each study site. At sites where the chemograph does not coincide with the hydrograph, as is the case at all five study sites, loads must be computed by multiplying 15-minute chemographs by 15-minute streamflow records and summing the products to compute mean daily loads. Therefore, dividing the mean daily load by the mean daily flow produces a flow-weighted daily concentration which is not equal to the mean daily concentration, unless concentration or flow are constant during the day.

When the USGS publishes mean daily sediment concentrations, these are time-averaged concentrations determined from a continuous chemograph (Porterfield, 1972). This time-averaging approach was used to compute mean daily, monthly, and annual concentrations at each study site. At the two Sangamon watershed sites, mean annual SSC ranged from 32 to 236 mg/L with an average value of 104 mg/L; at the three Spoon sites, mean annual SSC ranged from 37 to 193 mg/L with an average value of 83 mg/L. Mean annual TP concentrations are shown in Figure 6-5. Only during Water Years 2003, 2004, and 2009 were Cox Creek's mean annual TP concentrations comparable to the other study sites; during all other years Cox Creek's TP concentrations were much higher. Mean annual TP concentration at Cox Creek averaged 0.32 mg/L; at the other four study sites the annual TP concentrations averaged from 0.16 to 0.19 mg/L. Haw Creek had higher annual TP concentrations than the other two Spoon sites in 8 out of 10 years.

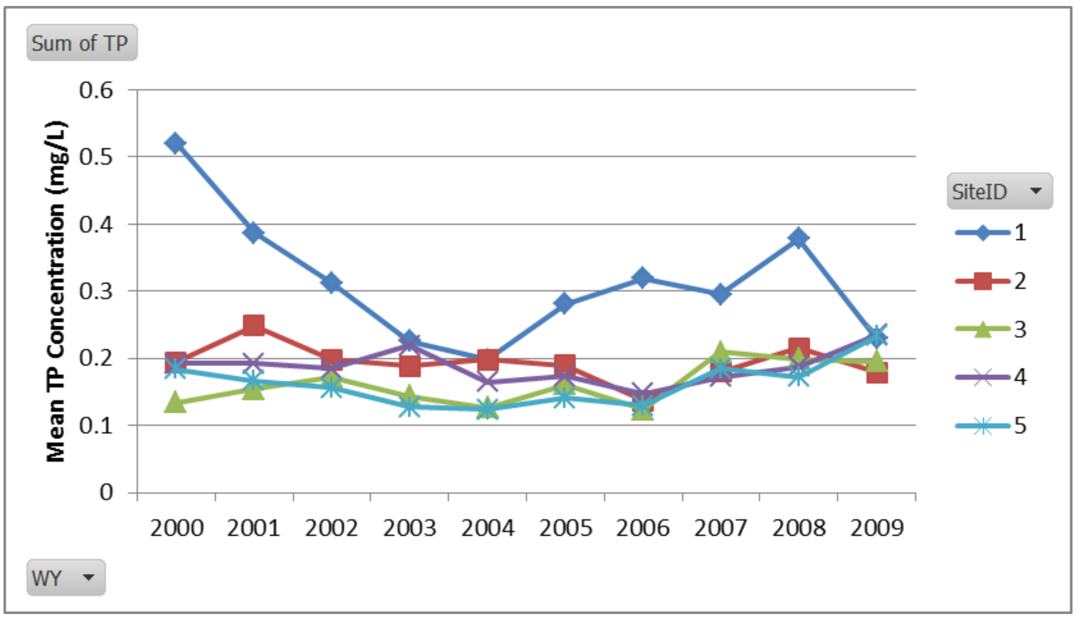

Figure 6-5. Mean annual TP concentrations at study sites, WY2000-2009

Annual median SS concentrations are presented in Table 6-11 as both time-weighted and flow-weighted values. The large differences between these two types of averages illustrate the high concentrations experienced during large runoff events. Annual median TP concentrations

(Table 6-12) illustrate a similar behavior. These two values should be interpreted appropriately to avoid misuse. The time-weighted concentrations describe what average stream concentrations are like on most days, but will dramatically under estimate concentrations during a runoff event. The flow-weighted concentrations, on the other hand, will over estimate concentrations in the stream on most days and are more appropriate to estimate stream loadings.

Table 6-11. Annual median sediment concentrations, mg/L

Site ID	Site Name	Time-weighted	Flow-weighted
1	Cox Creek	89	1,209
2	Panther Creek	76	1,154
3	North Creek	81	822
4	Haw Creek	71	469
5	Court Creek	88	813

Table 6-12. Annual median total phosphorus concentrations, mg/L

Site ID	Site Name	Time-weighted	Flow-weighted
1	Cox Creek	0.31	1.02
2	Panther Creek	0.19	0.77
3	North Creek	0.16	0.72
4	Haw Creek	0.19	0.72
5	Court Creek	0.16	0.87

Monthly FWCs for North Creek are good examples of the influence antecedent hydrologic conditions can have on the loading response of these stream sites. For example, 2008 produced the highest sediment FWC for the month of September, but its concentration was only triple September 2001's sediment FWC, yet September 2001's flow was one-tenth that of September 2008. Also at North Creek, even though October 2005 and 2008's monthly flows were comparable, October 2008 followed a wet September when perhaps sediment sources had been exhausted by multiple successive runoff events, so its sediment FWC of 35 mg/L and load of 16 tons were markedly lower than experienced in October 2005, a month which followed a very dry September. September 2005's sediment FWC was 261 mg/L and resulted in more sediment discharge (121 tons) than seen for similar flows three years earlier.

6.2 Summary and Conclusions

The sediment and total phosphorus loading rates determined at the five study sites were similar to those found in other studies in this region. Jacobson et al. (2011) estimated for Cass and Knox counties January-June TP yields of 0.28 to 0.39 kg/ha, which are quite a bit lower than the yields found in this study. However, they acknowledged that their estimates were lower than those found in other studies and may be due to their methods underestimating large flow events. According to Short (1999) annual mean TP concentrations during the period WY 1981-1996 in the Spoon and Sangamon watersheds were 0.15-0.30 mg/L and 0.30-0.45 mg/L, respectively. Annual total suspended solids concentrations during the same period in the Spoon and Sangamon were in the range of 150 to 308 mg/L and 100 to 150 mg/L, respectively. TP yields at the 5 study sites were also comparable to Short's yields for stations within the Spoon and Sangamon watersheds. The SS yields at my two Sangamon study sites were much higher than those reported by Short, but the three Spoon study sites were comparable. The nearest Sangamon watershed site in Short's study had a drainage area of more than 300 square miles and was located on a stream significantly impacted by a point source discharge. The annual median sediment yields at the five study sites were higher than median sediment yields reported for rural watersheds in Wisconsin, but the TP yields were actually comparable to several predominantly agricultural watersheds in Wisconsin (Corsi et al., 1997). The small watersheds in Wisconsin often had only a few years of record, and as shown in Section 6.1.1, annual variation can have a dramatic impact on yields. Annual sediment yields at the five study sites ranged from a minimum of 0.1 t/ac/yr at North Creek during WY 2004 to a maximum of 15.1 t/ac/yr at Cox Creek during WY 2008, and averaged 1.6 t/ac/yr at all study sites. Annual TP yields ranged from a minimum of 0.3 lb/ac/yr at Court Creek during WY 2006 to a maximum of 13.5 lb/ac/yr at Cox Creek during WY 2008, and averaged 2.3 lb/ac/yr.

The timing of sediment and phosphorus transport was also similar to, but perhaps even more dramatic than found in other studies. Royer et al. (2006) found that P export from predominantly agricultural watersheds in east-central Illinois largely occurred January-June. While I also found this generally to be true, over the whole ten-year study period the distribution of TP export was more even throughout the seasons at these study sites. This can most likely be attributed to the steeper slopes, larger percentage of forested lands in the watersheds, and higher rates of stream bank erosion found at my study sites. Extreme discharges at the five study sites accounted for greater than 90% of TP export, compared to 80% of TP export Royer et al. (2006) found transported during the top 10^{th} percentile of flows. This difference is most likely due to the smaller drainage areas of our study sites and the flashiness of our streams.

7. Conclusions and Further Work

Chapter 4 of this study evaluated eleven log-linear regression models for sediment concentration prediction. An 8-parameter model with terms for streamflow, seasonality, time trends, and a variable to describe the rate of change in flow was identified as the best performing model. Fourteen log-linear regression models were developed for total phosphorus concentration prediction, including three models which incorporated an instantaneous measure of suspended sediment concentration. The addition of SSC to the 8-parameter model described above was identified as the best performing model for TP prediction.

Chapter 5 of this study quantified the differences between four concentration estimation techniques (linear interpolation, regression models, regression models with error correction, and regression models with modified error correction) and their resulting load estimates. For both suspended sediment and total phosphorus, load calculations by error corrected regression models produced estimates that were the most precise and least biased of the four estimation techniques evaluated for these study sites. The method of error correction was not as critical as the act of error correction itself.

Chapter 6 of this study determined the proportion of sediment and phosphorus loadings as a function of time and flow. During the ten-year study period, 5% of the time accounted for approximately 50% of the flow, 91% of the total phosphorus load and more than 96% of the sediment load. On average, the 1-day maximum accounts for 10% of annual flow and more than 30% of annual sediment and phosphorus load.

Overall, the findings of this study support the conclusion that small, rural streams in western Illinois have a behavior that can be generalized for monitoring and estimating sediment and phosphorus loads in other small watersheds in Illinois. That behavior includes:

1. During a single storm event, sediment and phosphorus concentrations change by several orders of magnitude
2. High stream flashiness leads to sub-daily peaks in sediment and total phosphorus.
3. Most flow, sediment, and phosphorus is transported in an incredibly small proportion of time.

The importance of stream flashiness in sediment and phosphorus loadings in small, rural watersheds is summarized in Table 7-1 where annual median SS and TP yields at the five study sites were correlated with site characteristics. Neither annual median flow nor slope was strongly correlated with SS and TP yields. However, drainage area was inversely correlated, and stream flashiness was strongly correlated.

Table 7-1. Pearson's correlation coefficients showing relationships between annual median SS and TP yields and site characteristics for the five study sites

	SS Yield	TP Yield
Annual median flow	-0.25	0.03
15-min RBI	0.99	0.94
Daily RBI	0.87	0.73
Drainage area	-0.81	-0.66
Slope	0.27	-0.04

7.1 Recommendations

The results of this study have potential implications for various audiences: (1) researchers computing loads who are fortunate enough to have robust monitoring data, (2) researchers designing monitoring studies, especially those of short duration, and (3) design engineers and resource managers interested in reducing sediment and phosphorus loadings into their water body of concern.

7.1.1 Load Calculations at Existing Sites

The primary recommendation from this study is that error corrected regression models should be used to compute loads for sites that are intensively monitored with routine and storm event samples, especially if those sites are small, flashy streams where flows and/or concentrations change by several orders of magnitude in less than a day. These results should be applicable to any NPS pollutant even dissolved constituents like nitrate-N which may be diluted and tend to decrease in concentration during storm events. A secondary recommendation would be that error corrected regression models should not be used to compute loads at sites without adequate storm sample coverage.

On small, flashy streams I would advise against making the assumption that the concentration of sample collected during a runoff event is representative of mean daily conditions. This research clearly shows stream response on watersheds this size in this region is much too quick for that assumption to be valid. Individual sediment and phosphorus concentrations collected at Court Creek are plotted against the mean daily concentration computed for their date of collection in Figure 7-1 and Figure 7-2, respectively. While the assumption that an instantaneous sample is representative of that day's average conditions is appropriate at lower concentrations, the error in that assumption quickly grows to more than an order of magnitude at medium and high concentrations. For this reason, the use of regression models based on mean daily flows is discouraged for streams with high ratios of 15-min RBI to daily RBI, unless the mean daily concentration used in model development is time-weighted as recommended by Porterfield (1972).

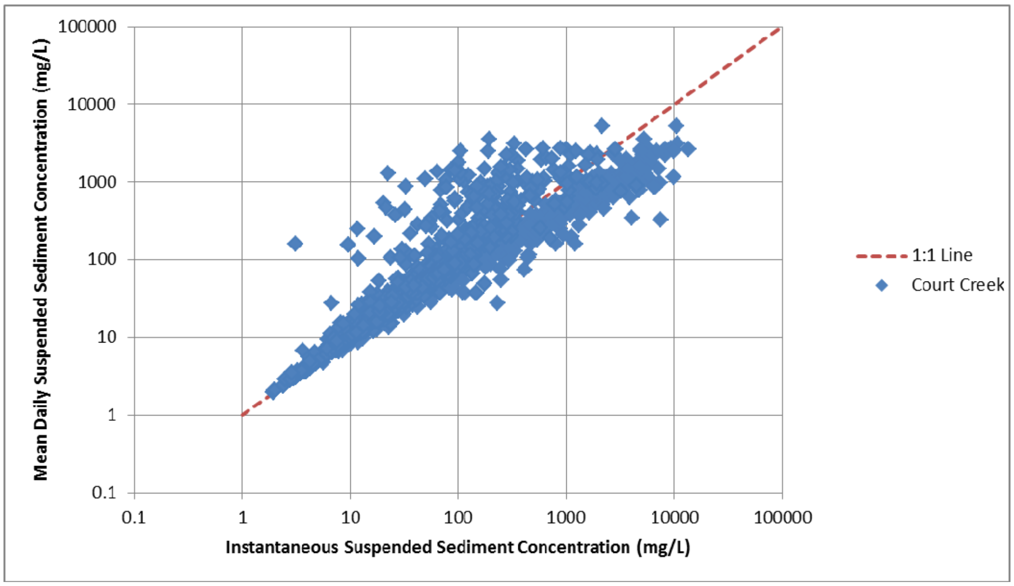

Figure 7-1. Relationship between instantaneous suspended sediment and mean daily suspended sediment concentrations at Court Creek

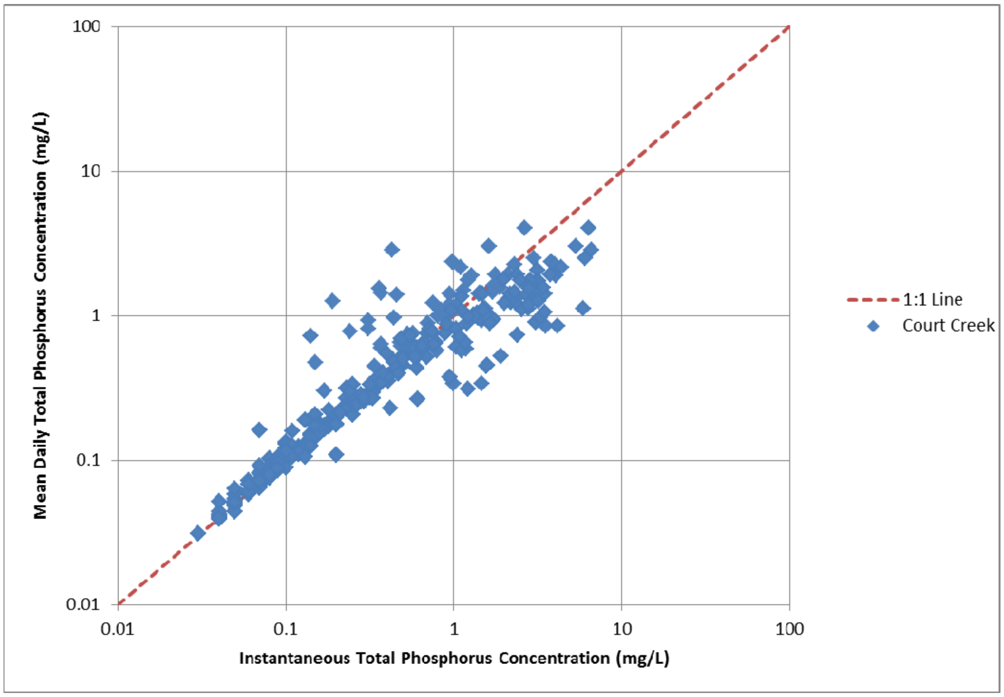

Figure 7-2. Relationship between instantaneous total phosphorus and mean daily total phosphorus concentrations at Court Creek

7.1.2 Future Monitoring Design

The result of this research with the largest implication for future monitoring design is the fact that it is absolutely critical to sample runoff events to accurately estimate loads. I would put forth that the risk of not sampling a runoff event is greater than the need to completely describe events, so I would recommend obtaining a couple samples for as many runoff events as possible rather than completely sampling a few events. In fact, I would suggest not trying to get complete coverage during storm events, but rather a minimum of at least two samples per event, one on the rising limb and one on the falling limb. If one of the monitoring objectives is model calibration, then it would definitely be beneficial to fully sample a few runoff events, but if the objective is to accurately account for sediment or phosphorus loads than I would recommend trying to obtain at least 1-2 samples for every event year-round.

Another recommendation from this research is using SSC as a surrogate to improve estimates of TP (or other particulate) concentrations and loads. If the monitoring study is interested in sediment and nutrient loads, the plan could include collection of sediment samples at a higher frequency than the other constituents in order to save on analytical costs while leveraging their correlation to improve nutrient load calculations. If this approach is followed, it is critical that the correlated constituent (e.g. TP) is sampled at the same time or as close as possible to the sediment sample.

7.1.3 BMP Design and Resource Management

Accurate load estimation is important because it can be used for not only identifying sources of pollutants prior to BMP development, but is also used after BMP implementation to document any water quality improvements. Any post-BMP monitoring conducted on small, rural streams to assess improvements in water quality must include targeted sampling of runoff events. Otherwise, it is entirely too easy to miss the periods contributing the greatest proportion of loads. Furthermore, BMPs designed to reduce loadings need to target high flow events. BMPs designed to reduce in-stream concentrations, however, can utilize smaller design storms or even baseflow dominated flows, as these are the flows a majority of the time. This study also highlights the importance of continuous flow records in any post-BMP monitoring. Obtaining measures of in-stream concentration without accompanying flow data is not sufficient; flow information is critical. Instantaneous measures of discharge at the time of sampling are not enough; in order to obtain information on the rate of change in flow and antecedent flow conditions, operation of a stream gage is required. The best way to truly document improvements is continued long-term monitoring.

7.2 Future Research

Based on this study, the following research questions are posed. Can this method of load calculation (RM10_CM) be used in short-term studies? Are the predictions in this study so successful because there were 10 years of data available to develop the regression models? How would the estimates be if only one or two years of observed data were used for regression model development? Would creating separate regression models for each water year actually improve the calculations, as Horowitz (2003) would suggest?

The results of this study are probably most appropriate in watersheds with significant non-point source pollution and high sediment erosion rates. Further research would be needed to determine whether these regression models would be applicable in areas with significant point source dischargers.

This study also only considered suspended sediment loads. No measures of bed load were available for analysis. Further research into particle size analysis of suspended sediments would be beneficial to understand the role of sand in the sediment loads at these sites. There are also potential implications as to the role of sand in phosphorus loads because of the relationship between stream P and bed sediments.

8. References

Aulenbach, B. T., and R. P. Hooper. 2006. The composite method: An improved method for stream-water solute load estimation. *Hydrological Processes* 20: 3029-3047.

Aulenbach, B. T., H. T. Buxton, W. A. Battaglin, and R. H. Coupe. 2007. Streamflow and nutrient fluxes of the Mississippi-Atchafalaya River Basin and subbasins for the period of record through 2005. Open-File Report OF-2007-1080. U.S. Geological Survey.

Baker, D. B., R. P. Richards, T. T. Loftus, and J. W. Kramer. 2004. A new flashiness index: Characteristics and applications to midwestern rivers and streams. *Journal of the American Water Resources Association* 40 (2): 503-522.

Birgand, F., C. Faucheux, G. Gruau, B. Augeard, F. Moatar, and P. Bordenave. 2010. Uncertainties in assessing annual nitrate loads and concentration indicators: Part 1. Impact of sampling frequency and load estimation algorithms. *Transactions of the ASABE* 53 (2): 437-446.

Bragg, H. M., and M. A. Uhrich. 2010. Suspended-sediment budget for the North Santiam River Basin, Oregon, water years 2005–08. Scientific Investigations Report 2010-5038.

Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference*. New York, NY: Springer-Verlag.

Cohn, T. A., D. L. Caulder, E. J. Gilroy, L. D. Zynjuk, and R. M. Summers. 1992. The validity of a simple statistical model for estimating fluvial constituent loads: An empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* 28 (9): 2353-2363.

Corsi, S. R., D. J. Graczyk, D. W. Owens, and R. T. Bannerman. 1997. Unit-area loads of suspended sediment, suspended solids, and total phosphorus from small watersheds in Wisconsin. Fact Sheet FS-195-97. U.S. Geological Survey.

Crowder, D. W., M. Demissie, and M. Markus. 2007. The accuracy of sediment loads when log-transformation produces nonlinear sediment load-discharge relationships. *Journal of Hydrology* 2007 (336): 250-268.

Demissie, M., L. Keefer, and J. Slowikowski. 2000. Quality Assurance Project Plan (QAPP) for CREP Monitoring Project. Unpublished. Champaign, IL: Illinois State Water Survey.

Demissie, M., L. Keefer, J. Slowikowski, A. Russell, T. Snider, and K. Stevenson. 2001. Sediment and nutrient monitoring at selected watersheds within the Illinois River watershed for evaluating the effectiveness of the Illinois River Conservation Reserve Enhancement Program (CREP). ISWS Contract Report CR 2001-12. Champaign, IL: Illinois State Water Survey.

Demissie, M., R. Xia, L. Keefer, and N. G. Bhowmik. 2003. Sediment budget of the Illinois River. *International Journal of Sediment Research* 18 (2): 305-313.

Edwards, T. K., and G. D. Glysson. 1999. Field methods for measurement of fluvial sediment. Techniques of Water Resources Investigations Book 3, Chapter C2. U.S. Geological Survey.

FISP. 1940. Field practice and equipment used in sampling suspended sediment. A Study of Methods Used in Measurement and Analysis of Sediment Loads in Streams Report No. 1. Iowa City, IA: Federal Interagency Sedimentation Project.

Glysson, G. D. 1987. Sediment-transport curves. Open-File Report 87-218. Reston, VA: U.S. Geological Survey.

Goolsby, D. A., W. A. Battaglin, G. B. Lawrence, R. S. Artz, B. T. Aulenbach, R. P. Hooper, D. R. Keeney, and G. J. Stensland. 1999. Flux and sources of nutrients in the Mississippi-Atchafalaya River Basin: Topic 3 Report for the integrated assessment on hypoxia in the Gulf of Mexico. NOAA Decision Analysis Series No. 17. Silver Spring, MD: NOAA Coastal Ocean Program.

Gray, J. R., G. D. Glysson, L. M. Turcios, and G. E. Schwarz. 2000. Comparability of suspended-sediment concentration data and total suspended solids data. Water Resources Investigations Report WRIR 00-4191. Reston, VA: U.S. Geological Survey.

Guo, Y., M. Markus, and M. Demissie. 2002. Uncertainty of nitrate-N load computations for agricultural watersheds. *Water Resources Research* 38 (10) 1185.

Haggard, B. E., T. S. Soerens, W. R. Green, and R. P. Richards. 2003. Using regression methods to estimate stream phosphorus loads at the Illinois River, Arkansas. *Applied Engineering in Agriculture* 19 (2): 187-194.

Helsel, D. R., and R. M. Hirsch. 2002. Statistical methods in water resources. Techniques of Water Resources Investigations Book 4, Chapter A3. U.S. Geological Survey.

Hirsch, R. M., D. L. Moyer, and S. A. Archfield. 2010. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs. *Journal of the American Water Resources Association* 46 (5): 857-880.

Horowitz, A. J. 2003. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. *Hydrological Processes* 17: 3387-3409.

IEPA. 2011. Illinois integrated water quality report and section 303(d) list - 2010 (draft). Springfield, IL: Illinois Environmental Protection Agency.

IEPA. 2012. Illinois integrated water quality report and section 303(d) list, 2012. Springfield, IL: Illinois Environmental Protection Agency.

Jacobson, L. M., M. B. David, and L. E. Drinkwater. 2011. A spatial analysis of phosphorus in the Mississippi River Basin. *Journal of Environmental Quality* 40 (3): 931-941.

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* 2007 (332): 241-258.

Knapp, H. V., and M. Markus. 2003. Evaluation of the Illinois streamflow gaging network. ISWS Contract Report CR 2003-05. Champaign, IL: Illinois State Water Survey.

Machesky, M. L., T. R. Holm, and J. A. Slowikowski. 2010. Phosphorus speciation in stream bed sediments from an agricultural watershed: Solid-phase associations and sorption behavior. *Aquatic Geochemistry* 16: 639-662.

Markus, M., and M. Demissie. 2006. Predictability of annual sediment loads based on flood events. *Journal of Hydrologic Engineering* 11 (4): 354-361.

McKallip, T. E., G. F. Koltun, J. R. Gray, and G. D. Glysson. 2001. GCLAS: A graphical constituent loading analysis system. *Seventh Federal Interagency Sedimentation Conference*. Reno, NV.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50 (3): 885-900.

Porterfield, G. 1972. Computation of fluvial-sediment discharge. Book 3, Chapter C3. U.S. Geological Survey.

Richards, R. P., D. B. Baker, J. W. Kramer, D. E. Ewing, B. J. Merryfield, and N. L. Miller. 2001. Storm discharge, loads, and average concentrations in northwest Ohio rivers, 1975-1995. *Journal of American Water Resources Association* 37 (2): 423-438.

Robertson, D. M. 2003. Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams. *Journal of the American Water Resources Association* 39 (5): 1281-1308.

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

Royer, T. V., M. B. David, and L. E. Gentry. 2006. Timing of riverine export of nitrate and phosphorus from agricultural watersheds in Illinois: Implications for reducing nutrient loading to the Mississippi River. *Environmental Science & Technology* 40 (13): 4126-4131.

Runkel, R. L., C. G. Crawford, and T. A. Cohn. 2004. Load Estimator (LOADEST): A Fortran Program for Estimating Constituent Loads in Streams and Rivers. USGS Techniques and Methods Book 4 Chapter A5. U.S. Geological Survey.

Short, M. B. 1999. Baseline loadings of nitrogen, phosphorus, and sediments from Illinois watersheds. IEPA/BOW/99-020. Illinois Environmental Protection Agency.

Simon, A., W. Dickerson, and A. Heins. 2004. Suspended-sediment transport rates at the 1.5-year recurrence interval for ecoregions of the United States: Transport conditions at the bankfull and effective discharge? *Geomorphology* 58: 243-262.

Slowikowski, J. A., A. M. Russell, K. E. Stevenson, and T. E. Snider. 2003. Instrumentation and sampling strategies for monitoring small watersheds in Illinois. *2003 ASAE Annual International Meeting*. ASAE Meeting Paper No. 032049. St. Joseph, MI: ASAE.

Sprague, L. A., R. M. Hirsch, and B. T. Aulenbach. 2011. Nitrate in the Mississippi River and its tributaries, 1980 to 2008: Are we making progress? *Environmental Science & Technology* 2011 (45): 7209-7216.

Terrio, P. J. 2006. Concentration, fluxes, and yields of nitrogen, phosphorus, and suspended sediment in the Illinois River Basin, 1996–2000. USGS Scientific Investigations Report 2006-5078. U.S. Geological Survey.

Vanni, M. J., W. H. Renwick, J. L. Headworth, J. D. Auch, and M. H. Schaus. 2001. Dissolved and particulate nutrient flux from three adjacent agricultural watersheds: A five-year study. *Biogeochemistry* 54: 85-114.

Verma, S., M. Markus, and R. A. Cooke. 2012. Development of error correction techniques for nitrate-N load estimation methods. *Journal of Hydrology* 432: 12-25.

Wang, P., and L. Linker. 2008. Improvement of regression simulation in fluvial sediment loads. *Journal of Hydraulic Engineering* 134 (10): 1527-1531.

Williamson, T. N., and C. G. Crawford. 2011. Estimation of suspended-sediment concentration from total suspended solids and turbidity data for Kentucky, 1978-1995. *Journal of the American Water Resources Association* 47 (4): 739-749.