# USING REGRESSION METHODS TO ESTIMATE STREAM PHOSPHORUS LOADS AT THE ILLINOIS RIVER, ARKANSAS

B. E. Haggard, T. S. Soerens, W. R. Green, R. P. Richards

**ABSTRACT.** *The development of total maximum daily loads (TMDLs) requires evaluating existing constituent loads in streams. Accurate estimates of constituent loads are needed to calibrate watershed and reservoir models for TMDL development. The best approach to estimate constituent loads is high frequency sampling, particularly during storm events, and mass integration of constituents passing a point in a stream. Most often, resources are limited and discrete water quality samples are collected on fixed intervals and sometimes supplemented with directed sampling during storm events. When resources are limited, mass integration is not an accurate means to determine constituent loads and other load estimation techniques such as regression models are used. The objective of this work was to determine a minimum number of water–quality samples needed to provide constituent concentration data adequate to estimate constituent loads at a large stream. Twenty sets of water quality samples with and without supplemental storm samples were randomly selected at various fixed intervals from a database at the Illinois River, northwest Arkansas. The random sets were used to estimate total phosphorus (TP) loads using regression models. The regression–based annual TP loads were compared to the integrated annual TP load estimated using all the data. At a minimum, monthly sampling plus supplemental storm samples (six samples per year) was needed to produce a root mean square error of less than 15%. Water quality samples should be collected at least semi–monthly (every 15 days) in studies less than two years if seasonal time factors are to be used in the regression models. Annual TP loads estimated from independently collected discrete water quality samples further demonstrated the utility of using regression models to estimate annual TP loads in this stream system.*

***Keywords.*** *Streams, Phosphorus loads, Phosphorus transport, Water quality monitoring.*

The Federal Clean Water Act Section 303(d) requires all states to identify streams and rivers that do not attain water quality standards and submit a list of these streams to the U.S. Environmental Protection Agency. This Act also specifies that the states must develop Total Maximum Daily Loads (TMDLs) or other watershed restoration approaches for streams listed. TMDLs are estimations of the amount of a constituent that streams can receive and still meet water quality standards. Thus, the first step in TMDL implementation is assessing existing constituent loads in streams. Agricultural engineers and scientists need accurate estimates of constituent loads for trend analysis, watershed and reservoir model calibration, and total maximum daily loads (TMDL) development. Often, long periods of hydrologic and water quality data are not available, and short–term studies, such as two years, are used to estimate constituent loads and calibrate watershed and reservoir models.

Constituent loads are a function of the volumetric rate of water passing a point in the stream and constituent concentration within the water. Several computer software programs are available to estimate constituent loads that pass a fixed point in a stream (e.g., GCLAS, U.S. Geological Survey, 2001; ESTIMATOR, Cohn et al., 1989, 1992; LOADEST2, Crawford, 1991, 1996). The data requirements are generally mean daily stream flow and constituent concentrations on days when water quality samples were collected. However, the frequency of water quality data collection required varies with the method used to estimate constituent loads, stream size and flashiness, and desired accuracy and precision. Often the data needs are greater than the information provided by typical (monthly, bimonthly, or quarterly) water quality sampling programs.

High–frequency water quality sampling, particularly during storm events, and using mass accumulation (integration) is the most accurate approach to estimate constituent loads if sufficient data are collected to describe the changes in water quality. Interpolation is generally used to estimate constituent concentration during time periods between water quality samples (Porterfield, 1972). Sufficient data often requires that many samples be collected during storm events to reflect the variability in constituent concentrations. And, samples collected at a single point within the stream cross section are calibrated to samples representing the integrated stream cross section (Ging, 1999; Martin et al., 1992), e.g., equal width and depth integrated (EWI) water quality

Article was submitted for review in July 2002; approved for publication by the Soil & Water Division of ASAE in January 2003.

Mention of a trade name, proprietary product or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply its approval to the exclusion of other products that may be suitable.

The authors are **Brian E. Haggard, ASAE Member Engineer,** Research Hydrologist, USDA–ARS Poultry Production and Product Safety Research Unit, Fayetteville, Arkansas; **Thomas S. Soerens,** Assistant Professor, Civil Engineering Department, University of Arkansas, Fayetteville, Arkansas; **W. Reed Green,** Water Quality Specialist, U.S. Geological Survey, Little Rock, Arkansas; **R. Peter Richards,** Director, Water Quality Laboratory, Heidelberg College, Tiffin, Ohio. **Corresponding author:** Brian E. Haggard, USDA–ARS–PPPSRU, 203 Engineering Hall, Fayetteville, AR 72701; phone: 479–575–2879; fax: 479–575–2846; e–mail: haggard@uark.edu.

samples (Edwards and Glysson, 1999). Constituent loads calculated by high frequency sampling and mass accumulation are often used as a reference to evaluate results from other methods such as ratio estimators or regression models (Robertson and Roerish, 1999; Preston et al., 1989; Richards and Holloway, 1987).

Regression methods use the relation between constituent concentration and stream flow to estimate missing daily concentrations. The analysis is usually performed following logarithmic transformation of concentration and flow. Regression methods often require less frequent water quality sampling than the integration method. The simple regression approach has been modified to account for nonlinearities, seasonal and long–term variability, censored data, logarithmic transformation biases, and residual serial correlation (Cohn, 1995). This more elaborate approach has been widely used, particularly as part of studies conducted by the U.S. Geological Survey (USGS) (Asbury and Oaksford, 1997; Pope and Milligan, 2000). Regression methods allow confidence limits to be placed on load estimates and are often used with relatively small datasets that have been assembled over many years.

Recently, water quality monitoring agencies have targeted storm events, collecting directed storm water quality samples as well as fixed interval samples. The timing or location of water quality samples collected during the storm event hydrograph may bias the annual TP load (Robertson and Roerish, 1999). For example, TP often displays hysteresis in streams that is the TP concentration may increase rapidly and peak on the rising limb of the storm and have a slower decrease in TP concentration on the falling limb (Richards et al., 2001; Richards and Holloway, 1987; Thomas, 1988). Robertson and Roerish (1999) suggested that in smaller streams additional storm samples resulted in a positive bias because the measured concentrations are typically greater than the average daily concentration. Furthermore, storm chasing by sampling crews as opposed to fixed stage sampling on the rising limb of the storm hydrograph resulted in the least biased load estimates of the storm sampling methods evaluated (Robertson and Roerish, 1999).

Our objectives were to compare regression–based total phosphorus (TP) load estimates to TP loads from mass accumulation in the Illinois River, Arkansas, and identify the number of water quality samples needed for use in regression models. This study compared the TP load calculated using mass accumulation of the entire database in 1998 (n = 449, Nelson and Soerens, 2000) with loads determined using the regression techniques and water quality samples randomly sub–sampled from the database (variable n) and with loads determined using the regression method and independently collected water quality data (n = 37). We also evaluated the use of seasonal time coefficients in regression models, effect of directed storm sampling, and the impact of targeting the storm hydrograph peak during storm chasing on the accuracy and precision of regression–based TP load estimates.

## Study Site Description

The study site is at an established U.S. Geological Survey streamflow and water quality monitoring station, the Illinois River, South of Siloam Springs, Arkansas (U.S. Geological Survey station number 07195430, Porter et al., 1999). The study site is in Benton County, Northwest Arkansas, about 8 km south of Siloam Springs on Arkansas Highway 59. The drainage basin area is approximately 1500 km$^2$. Overall, streamflow was generally greatest during winter and spring from 1997 through 2000, with some high flow events occurring in early summer. In general, low flows at this site occurred in summer and fall. Total stream flow was least in 1997 and greatest in 1999 (table 1). Average Base Flow Index (BFI, base flow proportion of total flow, Wahl and Wahl, 1995) was about 42%, ranging from 34 to 46%. About 40% of the streamflow occurred at a stage of about 1.5 m or less at the Illinois River during calendar year 1997–2000. The stage was less than 1.5 m on over 80% of the days during these years. Total phosphorus concentrations ranged from 0.06 to almost 1 mg/L and are related to streamflow. Total phosphorus concentrations decrease with increasing base flow but increase with increasing surface runoff discharge (Green and Haggard, 2001). The Arkansas Water Resources Center (AWRC) has estimated constituent loads at the Illinois River using high–frequency water quality sampling and mass accumulation from 1997 through 2000 (Nelson and Soerens, 2000, 2001; Soerens et al., 2000). Furthermore, the U.S. Geological Survey has also collected water quality samples and estimated nutrient loads using regression methods (Green and Haggard, 2001).

## Methods

### Water Quality Sub–Sample Sets

A detailed description of the water quality sampling protocol and laboratory analysis used by the AWRC can be found in Nelson and Soerens (2000). The AWRC collected grab samples from a single point during base flow conditions and collected samples from a single point using an automated sampler during storm events. From late September 1997 through May 1998, and during November and December 1998, water quality samples were collected every two days when the stage of the Illinois River was less than about 1.5 m or 18.4 m$^3$/s. When the stage was greater than 1.5 m, water quality samples were collected every 30 min during the rise of the hydrograph and every hour during the fall of the hydrograph (hereafter high–frequency sampling). During the other months involved in this investigation, discrete water quality samples were collected every two weeks and composite water quality samples were collected when the stage was greater than 1.5 m.

Because regression–based load estimates often use data assembled over multiple years, various sub–sample datasets were developed to mimic fixed–interval sampling and storm chasing from the 750 water quality samples collected by the

**Table 1. Streamflow characteristics of the Illinois River, south of Siloam Springs (U.S. Geological Survey Station No. 07195430) at Highway 59.**

| Year | Annual Mean Streamflow (m$^3$/s) | Annual Median Streamflow (m$^3$/s) | Total Streamflow (10$^6$ m$^3$) | Annual Peak Daily Streamflow (m$^3$/s) | Base Flow Index |
|---|---|---|---|---|---|
| 1997 | 14.1 | 8.1 | 444 | 643 | 0.44 |
| 1998 | 18.9 | 10.7 | 597 | 915 | 0.46 |
| 1999 | 20.9 | 11.3 | 659 | 830 | 0.44 |
| 2000 | 16.9 | 7.9 | 535 | 912 | 0.34 |
| Average | 17.7 | 9.5 | 559 | 825 | 0.42 |

AWRC during the period from late September 1997 through early June 1999. A random sampling date was selected during September 1997 using a random number generator, and then sampling dates were selected based on 15–, 30–, 45–, and 60–day intervals, corresponding to n = 39, 19, 14, and 10, respectively, from the original random sampling date in September or October 1997. If these intervals resulted in a date where the AWRC did not collect water quality data, then the sample collected closest to this date [in time] was selected. All samples were used in the development of these data sets, including individual storm and composite samples. A total of 20 sub–sample data sets were developed for each sampling interval (scheme), e.g. 20 sub–sample data sets for 15–day interval, 20 for 30–day, and so on. Because the AWRC collected water quality data less frequently when the stage of the Illinois River was less than about 1.5 m, the sub–sample data sets are not completely independent and could contain some of the same sample results within different datasets.

Storm events in this investigation are defined as times when the auto–sampler was triggered when the stage exceeded 1.5 m and high–frequency water quality sampling was used, not composite sampling. Sixteen individual storm events were sampled during this period. A random water quality sample was selected from each of the 16 storm events, and a total of 20 sub–sample sets were developed independently with 16 storm water quality samples. Water quality monitoring programs are not typically designed to collect discrete samples during each storm event when using grab–sampling protocols. Thus, 20 sub–samples sets from nine randomly chosen storm events were also produced representing six storm events per year sampling strategy. These 20 sub–sets of nine storm samples were limited to the upper 50% of the individual storm hydrograph because storm–sampling protocols often are directed toward near peak sampling. Twenty sub–sets of nine storms were also produced from the upper 25 and 10% of the storm hydrograph to ascertain any bias associated with targeting the peak of the storm event.

### Mass Accumulation

The INTEGRATOR software program (Richards, 2001) was used to estimate the TP load from the 449 samples collected by the AWRC during calendar year 1998 using numeric integration. The load associated with a sample interval is given by the average of the instantaneous flux (concentration × flow) at the time of the sample and the instantaneous flux at the time of the previous sample, multiplied by the time between samples. The annual load is the sum of all such loads. A limit may be placed on the length of time any sample is assumed to be valid. If any time interval exceeds this limit, then there is some unaccounted for time. The annual TP load can be adjusted for these unaccounted for time periods using the ratio of annual discharge to covered discharge, if desired.

This software program also provides an upper and lower bound to the annual load; however, these boundaries are not derived from statistical calculations. Rather, these bounds are the sums of the loads calculated using the larger and smaller of the fluxes that bound each time interval, rather than the average. This approach assumes that the fluxes increase or decrease monotonically within the interval between the samples. A modification is required at the local maxima and minima, because the assumption of monotonic change is false for one of the time intervals that surround such a point (unless the observed maximum or minimum is the true maximum or minimum, which is unlikely). The ad hoc approach used in this software assumes the true flux associated with a maximum is 125% of the observed maximum. Similarly, the true flux associated with a minimum is assumed to be 90% of the observed minimum.

### Regression Methods

This investigation used the relation between natural logarithm (ln) transformed concentration (C) or load (L) and daily mean discharge (Q), and seasonal factors (SIN and COS):

$$\ln(C) = \beta_0 + \beta_1 \ln(Q) + \beta_2 \,\mathrm{SIN}(2\pi T) + \beta_3 \,\mathrm{COS}(2\pi T) \quad (1)$$

$$\ln(L) = \beta_0 + \beta_1 \ln(Q) + \beta_2 \,\mathrm{SIN}(2\pi T) + \beta_3 \,\mathrm{COS}(2\pi T) \quad (2)$$

where C is concentration in mg/L, L is load in kg/d, Q is mean daily discharge $ft^3/s$, $\beta_0$ is the regression constant, T is time, and $\beta_1$, $\beta_2$, and $\beta_3$ are regression coefficients. Loads were estimated for each day using observed streamflow and summed to give annual loads. The ESTIMATOR software program was used to estimate annual TP loads in 1998 for each sub–set of data collected on assigned intervals with and without storm events. This software implements a minimum variance unbiased estimator to transform the results from ln space to real space (Cohn et al., 1989, 1992) and provides 95% confidence intervals of individual annual load estimates.

The full–scale regression model (eqs. 1 and 2) was used to determine annual TP loads using 15– and 30–day interval sub–sets with and without 16 storm events. This model also was used to evaluate changes in estimated annual TP loads from the 15–, 30–, 45–, and 60–day interval sub–sets with 9 and 16 storm events. However, seasonal factors were also omitted from the regression model when using only nine random storm samples. Thus, the simple relation between C or L and Q was also used in these cases:

$$\ln(C) = \beta_0 + \beta_1 \ln(Q) \quad (3)$$

$$\ln(L) = \beta_0 + \beta_1 \ln(Q) \quad (4)$$

The simple regression model (eqs. 3 and 4) was also used to determine the annual TP loads using 15–, 30–, 45– and 60–day interval sub–sets with nine random storm samples. This model was also used to assess biases associated with targeting the peak of the storm hydrograph in water quality monitoring programs, e.g. using the nine random storm events from the upper 50, 25, and 10% of the storm event hydrograph.

### Data Comparisons

We evaluated differences between the 1998 integrated annual TP load derived from the INTEGRATOR software and the load calculated by manual integration (Nelson and Soerens, 2000). The comparisons between annual TP loads estimated using INTEGRATOR and ESTIMATOR are presented as box plots. The difference between INTEGRATOR and the mean or median of the 20 regression–based TP load estimates was used to evaluate the accuracy (bias). The spread or standard deviation (SD) of the 20 regression–based

TP load estimates was used to evaluate the precision for each sampling protocol (Dolan et al., 1981; Preston et al., 1989). The accuracy and precision (bias and SD) were combined into the root mean square error (RMSE) where RMSE is defined by:

$$RMSE = (SD^2 + bias^2)^{0.5} \qquad (5)$$

In this study, the mean difference, median difference, median absolute difference, SD, and RMSE were used to evaluate the optimal water quality sampling strategies or number of water quality samples needed in regression techniques. The statistical values are calculated in kg/y and then presented in percent of the annual TP load estimate calculated via INTEGRATOR.

# Results and Discussion

## Mass Integration

Because of the occurrence of two–week sampling intervals when the stage was less than 1.5 m, there were times when loads were not calculated. The estimate from INTEGRATOR was corrected using the ratio of annual discharge to measured discharge used by the software program. The INTEGRATOR load was essentially equal to the load determined by manual integration. The INTEGRATOR load was 229,000 kg in 1998 whereas Nelson and Soerens (2000) reported 232,000 kg.

## Effects of Storm Samples

In order to demonstrate the significance of water quality sampling during storm events, 15– and 30–day datasets with and without directed storm chasing were used in regression models and the results compared to the integrated load in 1998 (table 2, fig. 1). Without directed storm samples, the 15– and 30–day datasets underestimated the annual TP load substantially. The mean and median of the 15–day predictions were 72 and 68% of the integrated annual TP load, respectively, and 30–day predictions were 64 and 58%, respectively. However, inclusion of directed storm–chasing data into the regression models substantially increased the accuracy, and the 15– and 30–day medians were within 10% of the integrated load. In fact, the 30–day median was within 1% of the integrated load. Some variation was observed in the

Table 2. Effect of additional, directed storm water quality samples on the accuracy and precision of regression–based annual phosphorus loads compared to the integrated 229,000 kg phosphorus load in 1998.[a]

| Fixed Interval Scheme (days) | Storm Samples (N)[b] | Total Samples (N) [c] | Mean Difference (%)[c] | Median Difference (%)[c] | Absolute Median Difference (%)[c] | SD (%)[c] | RMSE (%)[c] |
|---|---|---|---|---|---|---|---|
| 15 | – | 39 | –28 | –32 | 32 | 11 | 31 |
| 15 | 16 | 55 | –8 | –9 | 10 | 10 | 13 |
| 30 | – | 19 | –36 | –42 | 42 | 17 | 39 |
| 30 | 16 | 35 | –1 | –1 | 9 | 13 | 13 |

[a] The regression models were equations 1 and 2.
[b] The dash (–) suggests that storm chasing was not used; however, the fixed interval–sampling scheme may have included storm samples.
[c] The mean, median, and absolute median difference, standard deviation (SD) and root mean square error (RMSE) are defined as percent of the 229,000–kg phosphorus load determined using the INTEGRATOR software program. Storms samples were randomly sub–sampled from the entire storm hydrograph from 16 individual storm events.

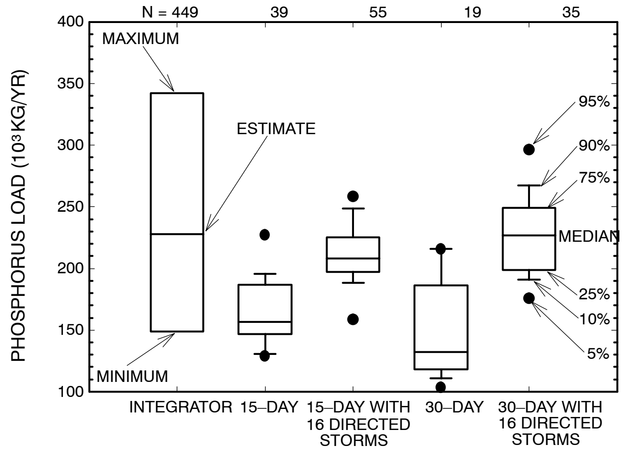


**Figure 1. Effect of additional, directed storm water quality samples on regression–based TP load estimates in the Illinois River, Arkansas. (The plot of INTEGRATOR represents the annual TP load estimate and the minimum and maximum bounds; regression–based load box plots represent the 10th, 25th, 50th, 75th, and 90th percentiles, and symbols represent the 5th and 95th percentiles. N represents the number of water quality observations used in integration or regression models.)**

regression–based predictions with and without directed storm samples, but the datasets that included directed storm samples were much better overall. Only load estimates greater than the 90th percentile using datasets not including directed storm samples approached the integrated load. These observations underscore the need to specifically target storm events in water quality monitoring programs.

The ability and resources of most agencies involved in water quality monitoring programs limit the number of samples that are collected. The ability to collect samples from all storms or 16 storms in 19 months is unlikely. However, the regression–based loads with 16 directed storm samples performed quite well regardless of the fixed sampling interval used (table 2 and 3). Water quality monitoring programs usually target a pre–determined number of storm events, not all storms. For example, the U.S. Geological Survey attempts to collect six storm events samples per year at several sites within the Illinois River Basin (Green and Haggard, 2001). The U.S. Geological Survey National Water quality Assessment (NAWQA) program was designed to collect four to eight storm samples per year (Hirsch et al., 1988). Thus, the remainder of comparisons between the regression–based and integrated annual TP loads uses only nine directed storm event samples per data set, or six storm samples per year.

All mean and median differences were within 10% of the integrated value and did not sequentially increase with specifically targeting the storm hydrograph peak, although all were slightly positively biased (table 4, fig. 2). The precision increased with limiting the portion of the storm hydrograph randomly sampled. However, reducing the portion for sub–sampling reduces the number of observations available for selection from each storm. The combination of accuracy and precision (RMSE) was within 15% of the integrated load for all portions of the storm hydrograph analysis (table 4). It is possible that in larger streams, such as the Illinois River, targeting the peak may be the best storm chasing technique as long as samples are randomly collected on either side, e.g., the rising or falling limb of the

Table 3. Statistical properties of 20 regression–based annual phosphorus loads from various water quality monitoring strategies using fixed–interval samples plus additional storm samples compared to the integrated 229,000–kg phosphorus load in 1998.[a]

| Fixed Interval Scheme (days) | Regression Equation Number | Storm Samples (N) | Storm Hydrograph (%) | Total Samples (N) | Mean Difference (%) | Median Difference (%) | Median Absolute Difference (%) | SD (%) | RMSE (%) |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 1,2 | 16 | 100 | 55 | –8 | –9 | 10 | 10 | 13 |
| 30 | 1,2 | 16 | 100 | 35 | –1 | –1 | 9 | 13 | 13 |
| 45 | 1,2 | 16 | 100 | 30 | 0 | –3 | 7 | 10 | 10 |
| 60 | 1,2 | 16 | 100 | 26 | 0 | –3 | 9 | 14 | 14 |
| 15 | 1,2 | 9 | 100 | 48 | –12 | –12 | 12 | 8 | 15 |
| 30 | 1,2 | 9 | 100 | 28 | –6 | –7 | 12 | 16 | 17 |
| 45 | 1,2 | 9 | 100 | 23 | 18 | 13 | 13 | 16 | 24 |
| 60 | 1,2 | 9 | 100 | 19 | 23 | 17 | 18 | 28 | 36 |
| 15 | 1,2 | 9 | Upper 50 | 48 | 2 | 3 | 7 | 10 | 11 |
| 30 | 1,2 | 9 | Upper 50 | 28 | 11 | 7 | 8 | 13 | 17 |
| 45 | 1,2 | 9 | Upper 50 | 23 | 15 | 12 | 12 | 14 | 21 |
| 60 | 1,2 | 9 | Upper 50 | 19 | 20 | 19 | 19 | 22 | 30 |
| 15 | 3,4 | 9 | Upper 50 | 48 | –5 | –5 | 8 | 9 | 10 |
| 30 | 3,4 | 9 | Upper 50 | 28 | 5 | 4 | 8 | 11 | 12 |
| 45 | 3,4 | 9 | Upper 50 | 23 | 8 | 3 | 10 | 14 | 16 |
| 60 | 3,4 | 9 | Upper 50 | 19 | 15 | 14 | 14 | 15 | 21 |
| 15 | 1,2 | 9 | Upper 25 | 48 | 6 | 5 | 10 | 10 | 11 |
| 30 | 1,2 | 9 | Upper 25 | 28 | 19 | 16 | 16 | 18 | 26 |
| 45 | 1,2 | 9 | Upper 25 | 23 | 42 | 36 | 36 | 24 | 48 |
| 60 | 1,2 | 9 | Upper 25 | 19 | 27 | 24 | 24 | 17 | 32 |
| 15 | 3,4 | 9 | Upper 25 | 48 | –2 | –3 | 5 | 8 | 8 |
| 30 | 3,4 | 9 | Upper 25 | 28 | 8 | 7 | 7 | 11 | 14 |
| 45 | 3,4 | 9 | Upper 25 | 23 | 18 | 15 | 15 | 14 | 23 |
| 60 | 3,4 | 9 | Upper 25 | 19 | 13 | 9 | 10 | 14 | 19 |

[a] Regression equation numbers refer to the equations in the text and are the model used to estimate phosphorus loads.

hydrograph. In contrast, Robertson and Roerish (1999) suggested storm chasing produced a positive bias and resulted in imprecise, overestimated loads.

The importance of storm events in nutrient transport and annual loads is substantial in streams and rivers. Greater than 85% of TP transported in the Illinois River occurs on days where surface runoff accounts for greater than 30% of total flow (Green and Haggard, 2001). This phenomenon is not localized and has been observed throughout the United States (Pionke et al., 1996; Richards et al., 2001). Almost half of TP transported in streams during storm events may be resuspended from bottom sediments (Svendsen et al. 1995). Release or resuspension of P associated with stream sediments in the Illinois River may be a critical source because

Table 4. Effect of targeting the storm hydrograph peak on accuracy and precision of regression–based annual phosphorus loads compared to the integrated 229,000–kg phosphorus load in 1998.

| Fixed Interval Scheme (days) | Storm Hydrograph (%) | Total Samples (N) | Mean Difference (%)[a] | Median Difference (%)[a] | Absolute Median Difference (%)[a] | SD (%)[a] | RMSE (%)[a] |
|---|---|---|---|---|---|---|---|
| 30 | 100[b] | 35 | –1 | –1 | 9 | 13 | 13 |
| 30 | Upper 50[c] | 28 | 5 | 4 | 8 | 11 | 12 |
| 30 | Upper 25[c] | 28 | 8 | 7 | 7 | 11 | 14 |
| 30 | Upper 10[c] | 28 | 8 | 8 | 8 | 10 | 13 |

[a] The mean, median, and absolute median difference, standard deviation (SD) and root mean square error (RMSE) are defined as percent of the 229,000–kg phosphorus load determined using the INTEGRATOR software program.
[b] Denotes regression equations 1 and 2 with 16 storm samples.
[c] Denotes regression equations 3 and 4 with 9 storm samples randomly sub–sampled from various portions of the storm hydrograph.

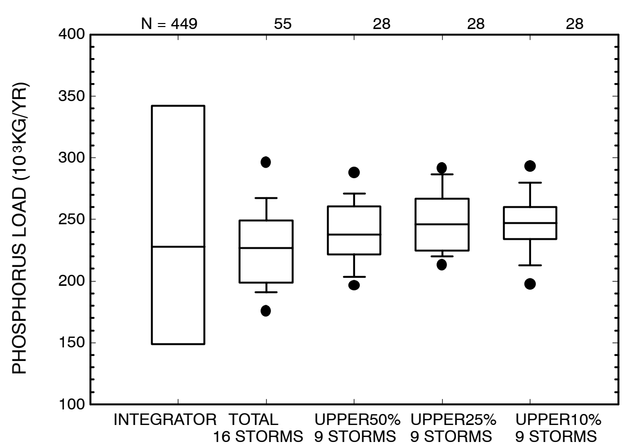


**Figure 2. Effect of chasing the storm hydrograph peak on regression–based annual TP load estimates in the Illinois River, Arkansas. (TOTAL 16 STORMS represents 16 individual storm samples collected from the entire storm hydrograph of each storm; UPPER50% 9 STORMS represents nine individual storm samples collected from the upper 50% of the storm hydrograph for each storm, and so on for 25 and 10%. N represents the number of water quality observations used in integration or regression models.)**

this stream receives P inputs from several wastewater treatment plants in the headwaters.

### Effect of Seasonal Time Coefficients

Regression models have the ability to include seasonal time coefficients in order to account for seasonal variations in TP concentrations observed in the residuals (Cohn, 1995); however, sufficient data needs to be collected to adequately describe these changes. It appears that when sampling all or 16 directed storm events over 19 months, the use of seasonal time coefficients in regression models functions well regardless of the fixed–interval scheme. But, when targeting a predetermined number of storm events and randomly selecting the storms, the use of seasonal time coefficients affected the accuracy and precision of the regression–based TP load estimates. In all comparisons (table 3), the use of seasonal time coefficients in regression models reduced the accuracy and precision, as the RMSE was consistently increased for all fixed sampling intervals. There was considerable variability in these reductions. The RMSE increased between 10 to 43% for all fixed intervals plus nine directed storm samples from the upper 50% of the storm hydrograph, and 50 to 113% for all fixed intervals plus nine storms from the upper 25%. In general, when using seasonal coefficients, the 15–day interval scheme plus nine directed storm samples performed better than any other fixed interval model with nine directed storm samples (the RMSE was 15% or less). This suggests that at a minimum, semi–monthly or 15–day fixed interval sampling is needed in approximately 19–month studies if seasonal time coefficients are to be included in the development of regression models to estimate TP loads. The alternative to this minimum sampling scheme was that all storm events should be sampled. Based on studies limited to less than two years in duration, the most important data requirement when using seasonal time factors is sufficient repetition of seasonal cycles.

### Water Quality Sampling Strategies

We must note that discussion with regard to water quality sampling strategies is limited to studies of two years or less, the time frame used in this analysis. Furthermore, this evaluation of the regression method is also limited to the Illinois River and other streams with similar TP concentration and flow relations. An analogous way to view these results is to focus on the number of water quality samples needed to develop regression equations and estimate annual TP loads. In short–term investigations, it appears that, at a minimum, monthly sampling plus six directed storm samples a year is required to adequately estimate annual TP loads in the Illinois River. The RMSE was within 15% of the integrated load when seasonal time coefficients were not used in the regression models and the upper half of the storm hydrograph was targeted (table 3). Several sampling schemes and regression models yielded accurate results. However, precision generally increased with an increasing number of observations or fixed–period sampling. Overall, the best water quality sampling strategy was 15–day fixed interval plus nine directed storm samples from the upper 25% of the storm hydrograph; the RMSE was only 8%.

In this study, the 28 observations were collected over a 19–month period; however, it may be feasible to collect this number of samples over a longer time frame. The collection of water quality data over time and with consistent methods allows the use of a rolling database. For example, if a water quality–monitoring agency is collecting bimonthly samples plus six storm event samples (n = 12 $y^{-1}$), then three years would be needed to cross the 28 observations threshold observed in this investigation. Thus, the regression model would use the past three years to estimate nutrient loads, e.g., the TP load in 2000 would be estimated using 1998 through 2000 data. Methods exist to determine if flow–weighted TP concentrations in streams are changing with respect to time, and these methods will identify the need to use a rolling database. The period of interest for calculated TP loads should coincide with the period of observation used to derive the regression equations.

The minimum number of observations suggested in this study is supported by independently collected EWI samples and annual TP loads estimated using ESTIMATOR at the Illinois River. Prior to October 1998, the U.S. Geological Survey collected water quality samples every other month, and after this date fixed–interval sampling was supplemented with about six storm water quality samples per year (Green and Haggard, 2001). Annual TP loads from 1997 through 2000 were based on 37 observations. The water quality samples were collected from the beginning of water year 1997 through the end of calendar year 2000 (all mean daily streamflow and TP concentration data are available at http://water.usgs.gov/ar/nwis). The regression–based annual TP load estimates from calendar year 1997 through 2000 compared well with the integrated estimates from Nelson and Soerens (2000, 2001) (fig. 3); the mean difference was less than 5%. In 1997, the annual TP loads estimated using the regression method with EWI samples was 17% greater than the integrated estimate (Nelson and Soerens, 2000). However, in 1998 through 2000 the EWI samples and regression models underestimated the annual TP load by an average of 12%. Overall, this comparison is quite remarkable considering that water quality samples were collected by different techniques, analyzed by different laboratories, and calculated by different methods.

While we only compared regression models, other investigators have suggested that ratio estimators were more

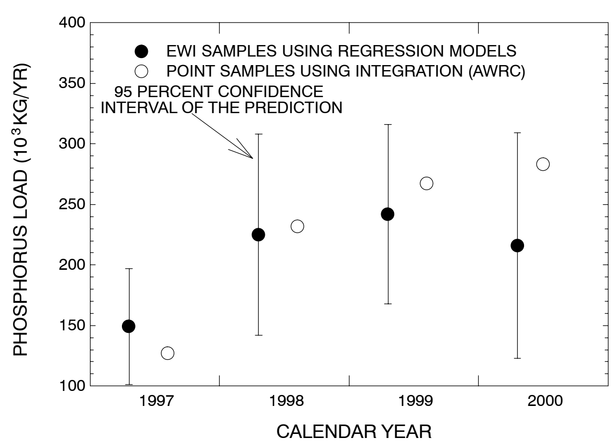


**Figure 3. Comparison between regression–based annual TP load estimates using EWI water quality samples collected by the U.S. Geological Survey and Arkansas Water Resources Center integrated (Nelson and Soerens, 2000, 2001) annual phosphorus load estimates using automated and point samples.**

robust to sources of error than other load estimation techniques such as averaging or regression models (Dolan et al., 1981; Preston et al., 1989). Furthermore, Richards and Holloway (1987) observed that loads based on flow–stratification using ratio estimators provided the best results. Similarly, Dolan et al. (1981) suggested that ratio estimators were better, even with concentrated sampling during storm events. The most recent study (Robertson and Roerish, 1999) evaluating the effect of sampling strategy on load estimation suggested that semi–monthly or 15–day sampling and regression methods were the least biased in small streams.

# CONCLUSIONS

This investigation focused on water quality samples collected over a period of 19 months. Thus, the identification of the most appropriate sampling strategy is based on 19 months of data collection, and this investigation suggested 28 water quality samples as the minimum number needed to adequately predict annual TP loads using regression models. However, the number of samples can be collected over a larger time frame than that used in this study, and the water quality samples should be collected during the period when loads will be estimated.

This investigation underscored the importance of specifically targeting storm events, and at the Illinois River and other larger streams, targeting the storm hydrograph peak may be the best approach to storm chasing. About semi–monthly or 15–day fixed interval sampling was needed if seasonal time coefficients are to be included in the development of regression models to estimate TP loads in our 19–month study. However, sufficient repetition of seasonal variations may be obtained with less frequent sampling for longer periods. Our sub–sampling comparison and the exceptional comparison between annual TP loads from 1997 through 2000 further validated the use of discrete sampling and regression models in annual TP load estimation, at least in streams similar to the one we studied.


### ACKNOWLEDGEMENTS

We are especially thankful for the assistance and information provided by the Arkansas Water Resources Center, and K. C. Davis, hydrologic technician, U.S. Geological Survey for numerous runs of the ESTIMATOR software program. The Arkansas Soil and Water Conservation Commission, Arkansas Water Resources Center, and the U.S. Geological Survey in Arkansas and Oklahoma provided funds for this investigation. This article benefited from comments by Dr. Peter Kleinman and three anonymous reviewers.


# REFERENCES:


Asbury, C. E., and E. T. Oaksford. 1997. A comparison of drainage basin nutrient inputs with instream nutrient loads for seven rivers in Georgia and Florida, 1986–1990. U.S. Geological Survey Water Resources Investigation Report 97–4006.

Cohn, T. A. 1995. Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers, U.S. National Report to the International Union Geodesy and Geophysics 1991–1994, Reviews of Geophysics, Supplement 33: 1117–1123.

Cohn, T. A., L. L. DeLong, E. J. Gilroy, R. M. Hirsh, and D. K. Wells. 1989. Estimating constituent loads. *Water Resources Research* 25(5): 937–942.

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

Crawford, C. G. 1991. Estimation of suspended–sediment rating curves and mean suspended–sediment loads. *J. of Hydrology* 129: 331–348.

_____. 1996. Estimating mean constituent loads in rivers by the rating–curve and flow–duration, rating–curve methods: Ph.D. dissertation, Indiana University, Bloomington, Ind.

Dolan, D. M., A. K. Yui, and R. D. Geist. 1981. Evaluation of river load estimation methods for total phosphorus. *J. of Great Lakes Research* 7(3): 207–214.

Edwards, T. K., and G. D. Glysson. 1999. Fields methods for measurement of fluvial sediment. *U.S. Geological Survey Techniques in Water Resource Investigations*, book 3, chapter C2.

Ging, P. B. 1999. Water quality assessment of South–central Texas – comparison of water quality in surface–water samples collected manually and by automated samples. U.S. Geological Survey Fact Sheet 172–99.

Green, W. R., and B. E. Haggard. 2001. Phosphorus and nitrogen concentrations and loads at Illinois River south of Siloam Springs, Arkansas, 1997–1999. U.S. Geological Survey Water Resources Investigation Report 01–4217.

Hirsch, R. M., W. M. Alley, and W. G. Wilber. 1988. Concepts for a National Water Quality Assessment Program. U.S. Geological Survey Circular 1021.

Martin, G. R., J. L. Smoot, and K. D. White. 1992. A comparison of surface–grab and cross sectionally integrated stream–water quality sampling methods. *Water Environment Research* 64(7): 866–876.

Nelson, M. A., and T. S. Soerens. 2000. Illinois River 1999 pollutant loads at Highway 59 bridge. Arkansas Water Resources Center Publication #MSC–279, University of Arkansas, Fayetteville, Ark.

_____. 2001. Illinois River 2000 pollutant loads at Highway 59 bridge. Arkansas Water Resources Center Publication #MSC–298, University of Arkansas, Fayetteville, Ark.

Pionke, H. B., W. J. Gburek, A. N. Sharpley, and R. R. Schnabel. 1996. Flow and nutrient export patterns for an agricultural hill–land watershed. *Water Resources Research* 32(6): 1795–1804.

Pope, L. M., and C. R. Milligan. 2000. Preliminary assessment of phosphorus transport in the Cheney Reservoir Watershed, South–Central Kansas, 1997–1998. U.S. Geological Survey Water Resources Investigation Report 00–4023.

Porter, J. E., D. A. Evans, and L. M. Remsing. 1999. Water Resources Data Arkansas Water Year 1999. U.S. Geological Survey Water–Data Report AR–99–1.

Porterfield, G. 1972. Computation of fluvial–sediment discharge. *U.S. Geological Survey Techniques of Water Resources Investigations*, book 3, chapter C3.

Preston, S. D., V. J. Bierman Jr., and S. E. Silliman. 1989. An evaluation of methods for the estimation of tributary mass loads. *Water Resources Research* 25(6): 1379–1389.

Richards, R. P. 2001. Written communication. User's guide for integrator: A program to calculate pollutant loads by numeric integration including upper and lower bound load estimates.

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. *J. of the American Water Resources Association* 37(2): 423–438.

Richards, R. P., and J. Holloway. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research* 23(10): 1939–1948.

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.

Soerens, T. S., M. A. Nelson, and J. Spooner. 2000. Investigation into the optimum sample number and timing for determining nutrient loads. Arkansas Water Resources Center Publication #PUB–182, University of Arkansas, Fayetteville, Ark.

Svendsen, L. M., B. Kronvang, P. Kristensen, and P. Graesbol. 1995. Dynamics of phosphorus compounds in a lowland river system: importance of retention and nonpoint sources. *Hydrological Processes* 9: 119–142.

Thomas, R. B. 1988. Monitoring baseline suspended sediment in forested basins: The effects of sampling on suspended sediment rating curves. *Hydrologic Science* 33: 499–514.

U.S. Geological Survey. 2001. Water resources application software, GCLAS, graphical constituent loading analysis system. Accessed 5 Dec. 2000. http://hassrvares.erusgs.gov/gclas.html.

Wahl, K. L, and T. L. Wahl. 1995. Determining the Flow of Comal Springs at New Braunfels, Texas: *Texas Water'95*, 77–86. Reston, Va.: ASCE.