

Mean Annual Acidity Load: A Performance Measure to Evaluate Acid Mine Drainage Remediation

Mary W. Stoertz and Douglas H. Green

Dept. of Geological Sciences, Ohio University, Athens, OH 45701, Stoertz@ohio.edu

ABSTRACT

Mean annual acidity load is proposed as a method for measuring performance of acid mine drainage remediation projects in terms of water quality. The method is not a substitute for biocriteria in measuring attainment of aquatic life uses. It allows public managers to assess and report the effectiveness of source control, passive treatment and active treatment in alleviating acidity loads, and also provides a basis for comparing cost-effectiveness of alternatives. The mean annual load is a representative measure of water quality because it integrates all probable daily loads over a typical year. The proposed method requires minimal data: drainage area, acidity and discharge, for high- and low-flow conditions, before and after treatment. Acidity load (acidity x discharge) is graphed against normalized discharge (normalized to mean annual discharge) in log space using Excel. Linear regressions are performed separately on pre-treatment and post-treatment data. The difference between the two regressions at normalized mean annual discharge is the change in the mean annual acidity load. Mean annual discharge can be estimated with high confidence from drainage area, for southeast Ohio basins with record lengths exceeding 25 years. The functional relationship between acidity and discharge for sites with sufficiently large data sets indicates that the governing physical model is dilution, as evidenced by a power-law relationship between acidity and discharge. In the limiting case of a constant acidity source, the exponent would be unity. In all tested cases, the exponent is less than unity, implying a non-constant acidity source in which acidity generation is greater at low flow.

Key Words: acid mine drainage, acidity, load, TMDL, water quality, discharge, reclamation, non-point source, dilution, flushing

INTRODUCTION

Much remains to be done in the restoration of rivers. Even though public managers and watershed groups are energetically undertaking stream restoration across the United States, over 40% of assessed U.S. waters still do not meet water quality standards states have set for them. These polluted lakes and streams are close to home: They lie within 10 miles of 218 million Americans (EPA, 2003).

States are required to maintain a list of waterbodies that fail to meet designated uses, and are required to develop plans for alleviating contaminant loads to these waterbodies. EPA's Total Maximum Daily Load (TMDL) program, promulgated under the authority of Section 303(d) of the 1972 Clean Water Act and amended in 1992, recognizes that chemical *load*, and not just *concentration*, is important in measuring the impact of a contaminant source. Load is the product of concentration (usually in mg/l) and discharge (usually in cfs), and is expressed in units of mass/time (for example, tons/day).

As projects to reduce non-point source loads are undertaken and completed, the National Research Council (2001) recommends that states make a periodic assessment of the waterbodies to determine if designated uses have been attained. Limited budgets prevent states from periodically monitoring *all* indicators even on major rivers, so any method that uses scarce data yet still provides reasonable estimates of loads and load reductions is valuable (NRC, 2001). The Clean Water Act, moreover, requires that water quality be maintained once the designated use is attained, so the monitoring proposition is long term.

Measuring progress in non-point-source load reduction is a matter of practical importance. Tools are needed to support decision-making in stream restoration. Restoration groups need to know if stressors are abating. They also need to report their progress to grantors, in particular the Environmental Protection Agency, which administers the Section 319 (Non-Point Source) program. Ohio EPA, for example, has a target of achieving aquatic life goals in 80% of Ohio streams by 2010 (www.epa.state.oh.us/dsw). Finally, as work continues, treatment assessment is needed to compare the efficiencies of various restoration strategies.

A variety of tools might be used to measure progress, including administrative indicators (e.g., permits issued), stressor indicators (e.g., effluent reduction), exposure indicators (e.g., decrease in in-stream pollutant concentrations), and response indicators (e.g., biometrics) (Karr and Yoder, 2004). Of these, only biometrics directly measure the end outcome, and have been

called the “gold standard endpoint” (Karr and Yoder, 2004). Chemical criteria are not reliable indicators of biological condition, so are not substitutes for biometrics. In a study by Davis et al. (1996), 25% of river miles were found degraded based on chemical criteria, but twice as many miles (50%) were degraded based on biological criteria.

On the other hand, stressor and exposure indicators have an important role because they correlate more directly with restoration activities, enabling managers to decide among treatment alternatives as restoration proceeds. For example, selection of active vs. passive treatment systems for acid mine drainage (Figure 1) might be decided based on stressor loading reduction and cost effectiveness. Biological indicators such as the Index of Biotic Integrity (IBI; Karr, 1981) and the Invertebrate Community Index (ICI; OEPA, 1987) directly measure attainment of aquatic life use, but may take time to respond to restoration activities that correct a water-quality deficiency. Furthermore, it is not clear how improvement of IBI or ICI correlates with specific activities (e.g. liming, fencing cattle, or eliminating a point source). Stressor and exposure indicators, while valuable tools, cannot replace biological criteria, which are the final measure of whether restoration endpoints have been reached (Karr and Yoder, 2004).

In this paper, we present a method that can be described as a stressor indicator, because it estimates the mean annual load of acidity in acid-mine-drainage (AMD) impaired streams, as well as changes in that load in response to treatment.



Figure 1. Strategic treatment of AMD by a suite of methods such as active lime sand dosing operation (left) or passive treatment wetlands (right), demands tools for performance measurement.

Acidity is an appropriate criterion to assess partial recovery of a mining-impaired stream toward support of its designated aquatic life use (e.g., Warmwater Habitat; OEPA, 1988), because it has a direct causal link to mining, responds rapidly to remediation, allows comparison

of alternative treatments (Ziemkiewicz et al., 2003), and is easy to measure via titration or to estimate from pH and mineral acidity (Skousen and Ziemkiewicz, 1996).

In streams impaired by acid mine drainage (AMD), the primary stressor is water quality. In fact, physical aquatic habitat may be excellent apart from water chemistry (Stoertz et al., 2002). Acidity derives from the main characteristics of AMD, namely low pH and high concentrations of iron, manganese and aluminum (Skousen and Ziemkiewicz, 1996). These constituents are known to limit fish and macroinvertebrate communities (Lacroix, 1987). Acidity can be measured directly by titration, but it can also be estimated from pH and mineral acidity (Equation 1).

$$\hat{a} = 50 (2[Fe^{+2}]/56 + 3[Fe^{+3}]/56 + 3[Al^{+3}]/27 + 2[Mn^{+2}]/55 + 1000(10^{-pH})) \quad (\text{Eq. 1})$$

where \hat{a} = estimated acidity (mg/l), brackets denote concentrations (mg/l), denominators are molecular weight, multipliers are species charge, and the 50 factor transforms milliequivalents (meq) of acidity into mg/l CaCO₃ equivalent.

With unlimited resources, chemical quality would be measured continuously, or at least daily. This ideal is seldom achieved, especially in citizen-based watershed groups. The challenge for most projects is to interpret trends in scattered water-quality measurements, in order to make decisions, justify expenditure of resources, and report changes.

Note: We use units of cfs for discharge, square miles for drainage area, and mg/l for acidity, because public managers typically have access to USGS discharge data in cfs, USGS topographic maps in miles, and lab analytical reports in mg/l. The graphs in this paper were made using Microsoft Excel, which allows insertion of trendlines and their equations, important for the method. Any software with this capability is suitable.

A typical data set collected at a sampling station (Figure 2) shows acidity over an eight-year period, with a dozen samples (n=12) during that time. Installation of an open limestone channel was completed in June 2001. The mean acidity has decreased from 1050 to 250 mg/l, a statistically significant decrease of 76%. It is clear from Figure 2 that the acidity is flow-dependent, so a strict average may be misleading.

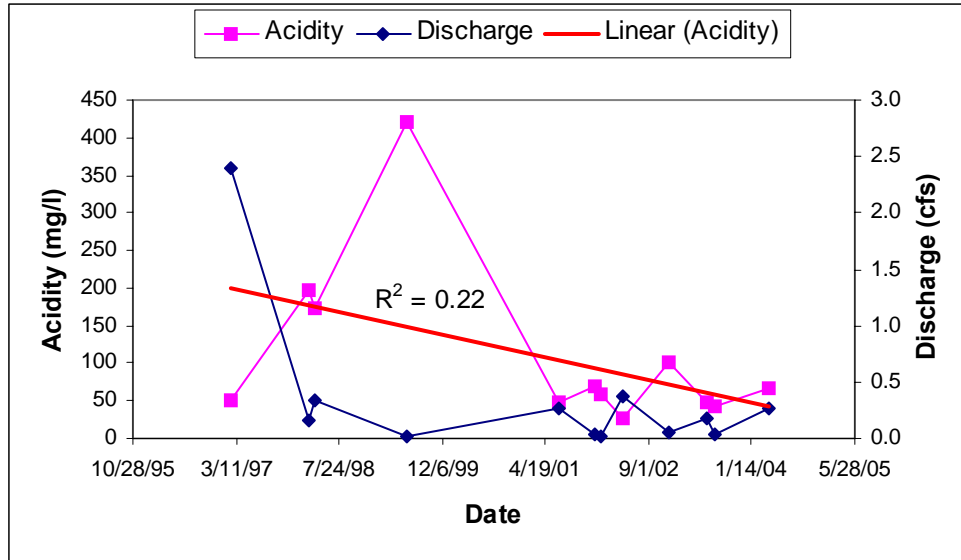


Figure 2. A typical data set shows a trend in acidity from 1997 (pre-treatment) to 2004 (post-treatment). However, the data are variable and flow-dependent, so that a strict average is misleading.

Methods are available for quantifying trends in environmental data, with various limitations. Time-series autocorrelation methods such as the Box-Jenkins method (Montgomery and Johnson, 1976) require 50-100 evenly spaced measurements. Non-parametric methods are an alternative, requiring no assumption of a normal distribution, and allowing missing data or unevenly spaced measurements. Examples of non-parametric trend estimators are Sen's test (Sen, 1968), the Kendall test (Mann, 1945; Kendall, 1975), and the seasonal Kendall test (Smith et al., 1982), which allows for seasonally variable data (Gibbons, 1994). Unfortunately, if the discharge is an important source of variability, the water quality time series will be serially correlated, invalidating tests of the significance of trends (Darken et al., 2002). In the example (Figure 2), discharge is clearly a source of acidity variability: high acidities are measured during low discharge, and vice versa. High flows are critical because of their associated high loads. However, monitoring is seldom conducted for non-point pollution sources during these important high-flow, high-rainfall events (NRC, 2001).

Several approaches have been used to explicitly incorporate flow to correct for the variability introduced by discharge, and many of them focus on load rather than concentration. Likens et al. (1977) averaged concentrations at the beginning and end of each sampling interval, multiplying the average concentration by the total discharge during that interval to obtain period-weighted loads. Johnson (1979) performed linear regressions of discharge and concentration for each month to estimate monthly loads from representative monthly discharge, and summed these

to get annual loads. They refined their approach by dividing the range of discharges into quartiles, each representing 25% of mean annual discharge, and selecting an equal number of existing samples from each quartile before performing the regression. In this way, important high-flow data are well represented and the regression is not biased by frequent low-flow samples. The flow-quartile approach is like sampling a stream in proportion to discharge, but has the practical disadvantage of requiring a large number of samples. Stednick and Roig (1989) applied the discharge-interval concept to design strategic field sampling, requiring far fewer samples. Hirsch et al. (1981) corrected constituent concentrations for flow by estimating acidity as a function of discharge. They subtracted this estimated acidity for each data point from the measured acidity to get a flow-adjusted data set that they analyzed using the Seasonal Kendall Test for trend. Dann et al. (1986) compared several methods and concluded that the Likens et al. (1977) period-weighted load gave the best estimates of the true load. Of course, frequent sampling is implied, and disproportionate representation of either baseflow or high-flow samples will still bias the estimated load. In cases where frequent or continuous discharge data are available, flow duration curves can be constructed and used, together with the discharge-concentration regression, to generate load rate duration curves (LRDC; Bonta and Dick, 2003). From the load duration curve, an average constituent load can be computed. Investigators have used flow duration curves to explicitly incorporate this variability into evaluations of the risk of load rates exceeding critical levels (Bonta and Cleland, 2003), and applied the method to estimating the impact of mining and reclamation on load rates to Ohio watersheds (Bonta and Dick, 2003).

Zetterqvist (1991) modeled phosphorous concentration ([P]) as a function of several causal variables, among them streamflow, considering it explicitly as a series influencing the [P] series, and related the streamflow series to water quality through a backward-shift operator that also allows for a lag time. She found a significant dilution relationship between [P] and discharge, consistent with a relatively constant P point source from sewage treatment plants. However, after modeling [P], much of the original variability remained as “noise.” Zetterqvist attributed the high noise level in part to a complex relationship between P and discharge not satisfactorily described by a linear transfer function model.

This paper proposes a load-based method in which pre-treatment acidity loads are regressed against discharge separately from post-treatment data, in log-log space. The

transformed functions are shown to be linear because acidity behaves according to a power law, implying that flushing and dilution are flow-dependent mechanisms controlling acidity. Comparing pre- and post-treatment acidity loads at the mean annual discharge yields a performance measure: the change in the mean annual acidity load. Data from AMD-impacted streams in the Western Allegheny Plateau were used to develop and test the method. The method may apply to other ecoregions and other pollutants, provided a power-law (i.e., dilution) relationship between concentration and discharge prevails. Provided high- and low-flow events are represented, mean annual acidity load can be estimated with high confidence, even with sparse data.

Antonopoulos *et al.* (2001) used a variant of the method developed in this paper, and similarly found that a load vs. discharge regression gave better results than a concentration-discharge regression. Ormsbee *et al.* (2004) used a similar approach to calculate a target load (TMDL) for hydrogen ions. The TMDL in that study was the hydrogen ion load that occurred at a threshold flow below which pH would violate a minimum-pH criterion based on aquatic life tolerance. The lowest average annual discharge of the most recent 10 years was used as the threshold, as determined by drainage area and a discharge-area relationship derived from USGS gages. Their method incorporated a margin of error in the hydrogen ion load, and took into consideration low concentrating flows. It was appropriate as an enforcement goal in active mining areas. Notably, EPA Region IV approved the Ormsbee *et al.* (2004) method as an acceptable TMDL protocol for pH.

METHOD DEVELOPMENT

The mean annual acidity loading method assumes that the acidity vs. discharge relationship follows a certain physical model such that the transformed variables are linearly related and are amenable to linear regression methods. In this section, hypothetical cases are used to show possible relationships and to interpret them physically and mathematically. Then, once the hypothetical cases are presented and interpreted, the method is applied to water-quality data from streams in the Western Allegheny Plateau to verify that real-world cases are consistent with the proposed dilution model.

Other researchers have observed several patterns or models for solute concentrations in streams. Concentrations may decrease in inverse proportion to discharge if the main source of

the constituent has a relatively constant source such as a point source or natural aquifer baseflow. Concentrations at high discharges may also be either higher or lower than expected for the pure-dilution case. Lewis and Grant (1979) describe these two cases as the “purging effect” and the “sparing effect.” Concentrations may even increase with discharge by storm “washoff” or by suspended-state transport (Hirsch et al., 1981). The different cases result because stream water is a mixture of separate water sources (e.g., Johnson and Likens, 1969). Strong correlations between concentration and discharge need not be dilution. Bonta and Cleland (2003) observed positive correlation between [Mn] and discharge during a period of watershed disturbance, interpreted as exposure of more oxidation products to transporting water. In underground mines, concentrations show a slight response to a “spring flush” but on the whole are relatively constant and independent of flow (López and Stoertz, 2001; Pigati, 1997; Stachler, 1997). In such cases it is appropriate to use mean acidity concentrations, simply averaging all concentrations regardless of flow.

The functional dependence of concentration on discharge is important. The proposed method assumes that concentration behaves according to a dilution model, which encompasses the case of independence of concentration and discharge, as has been observed in underground mine discharges. Other models are possible. Uniform concentrations are consistent with a mechanism of chemical equilibrium control of concentrations (López and Stoertz, 2001). AMD may, on the other hand, exhibit a “spring flush” phenomenon (Smith and Shumate, 1971), as stored oxidation reaction products are washed from the mine walls or refuse pores in the typically heavy recharge events of the spring, causing higher concentrations. These events may appear as outliers, but they are important components of the annual loading. A system that exhibits a spring flush may not be amenable to analysis by the proposed method.

Lewis and Grant (1979) explored relationships among concentration, discharge and load (or yield) of dissolved substances in a Colorado mountain stream, and observed three general patterns that they interpreted mechanistically. Discharge either decreased with increasing flow according to a dilution mechanism, or was independent of flow, or increased with flow by a flushing mechanism that mobilizes substances more effectively at high flow rates. Loads, similarly, were observed to a) increase at the same rate as discharge, b) increase more slowly than discharge by a “sparing” mechanism, or c) increase faster than discharge, by a “purging” mechanism. They explained the different mechanisms by invoking dual water sources following

the lead of Johnson et al. (1969): soil water and precipitation. Stream water can be seen as a mixture of the source waters, which in some cases may also react. Analogously, water in AMD-impaired streams may be a mixture of “baseflow” from deep mines and refuse piles, and recent precipitation or runoff. Smith, Hirsch and Slack (1982) examined trends in total P at over 300 NASQAN (National Stream Quality Accounting Network) stations, and found 204 significant regressions between P concentration and stream discharge. The most common pattern was hyperbolic (62% of cases), and other patterns were linear (20%), inverse (14%), and logarithmic (4%). Increasing concentrations with flow were interpreted as erosion and transport of P at high flows; decreasing concentrations with flow were interpreted as dilution of point-source contributions or subsurface dissolved sources (i.e., baseflow).

When a sampling station is downstream of a large reservoir, concentration may be independent of discharge because the water flows from a well-mixed source (Liebermann et al., 1987). Deep mine pools may act as reservoirs, with relatively uniform concentrations (López and Stoertz, 2001). In stream headwaters where deep mine discharge dominates the flow, concentrations may be independent of discharge. On the other hand, mine discharge may constitute a point source or baseflow source that is diluted by surface water. Mine discharges show strong seasonality in flow, but the flow is controlled by evapotranspirative demand on an annual cycle rather than by precipitation (Stoertz et al., 2001; Stoertz et al., 2004).

In the case of AMD discussed here, stream water can be viewed as a mixture of mine baseflow, runoff, and unsaturated seepage from coal refuse piles. Salient features of the acidity-producing systems are the rate of acidity production, which depends on availability of oxygen for the manifestation of mineral acidity, the presence of neutralizing alkalinity, and storage of water and reaction products. For example, a deep mine may be fully flooded and produce no acidity, or it may have a seasonally fluctuating water level that washes oxidation products from the mine walls, storing them in the acidic mine pool to be released slowly as baseflow (López and Stoertz, 2001; Stoertz et al., 2001). As another example, a refuse pile may experience rapid internal exothermic oxidation exacerbated by convective airflow through the hot pile, with acidity flushed out during a rainstorm. Mining or reclamation, which change these features, can alter trends such that the relation between discharge and concentration, and even the sign of the regression, may change (Bonta and Dick, 2003).

The three cases, dilution, flushing (purging), and sparing, can be understood by viewing hypothetical cases. Dilution of a constant source of acidity (a) follows a power law (Figure 3), in which doubling discharge (Q) halves acidity. Equation 2 governs this case, where c is a constant, representing the source load at unit discharge.

$$a = c/Q^m \quad (\text{Eq. 2})$$

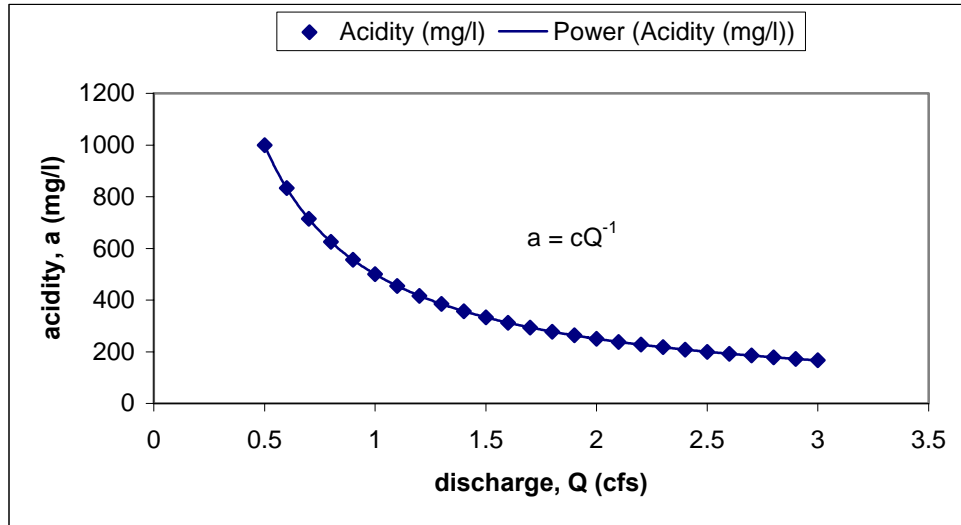


Figure 3. Dilution of a constant source of a conservative solute follows a power law. In this case, an arbitrary source load of 500 cfs·mg/l is used to illustrate the dilution principle, in which doubling discharge halves concentration. At any discharge, the load remains constant.

In pure dilution, the exponent $m=1$; in flushing, $m<1$; in sparing behavior, $m>1$.

Transforming the exponent makes it physically intuitive: By defining a flushing factor F , where $F=1-m$, then $F>0$ for flushing, $F=0$ for pure dilution, and $F<0$ for sparing behavior. All three cases are illustrated for a hypothetical case (

Figure 4), along with a partial flushing case, which is commonly observed in the field so it is presented here. Flushing occurs ($F>0$) when the contaminant load increases with discharge.

Theoretically, the concentration can be constant (load increases in proportion to flow; $F=1$), if the streamflow is derived from displacement of water from a uniformly polluted source such as a partially flooded mine, for example. Most commonly, partial flushing occurs: load increases at a lesser rate than discharge ($0<F<1$). Occasionally, load increases *faster* than discharge ($F>1$), due to washing of reaction products from mine walls or the unsaturated zone of a refuse pile. Sparing behavior can be interpreted as shutting down of reactions at high flow, perhaps due to flooding of reactive sites. Dilution behavior can be interpreted as a constant source.

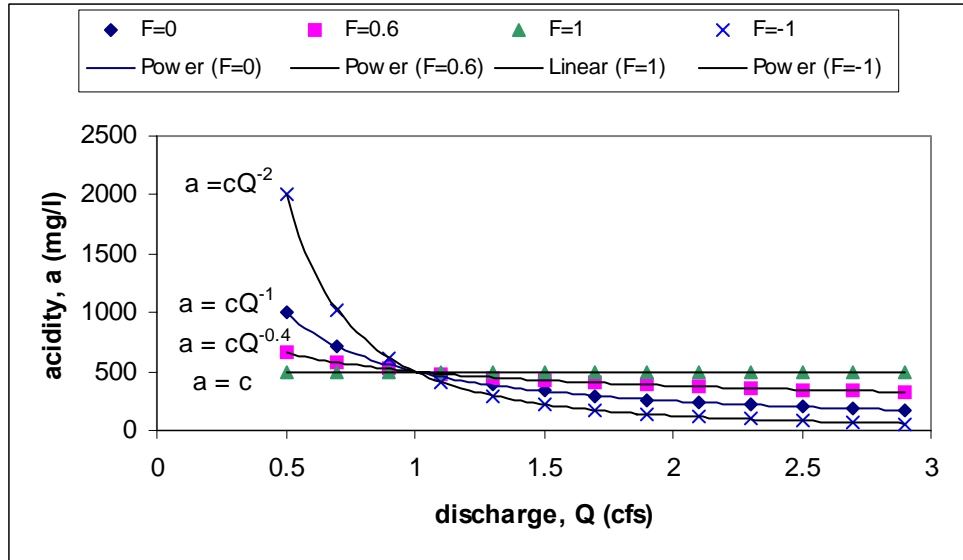


Figure 4. Hypothetical cases of pure dilution ($F=0$), pure flushing ($F=1$), partial flushing ($0 < F < 1$), and sparing ($F < 0$). Only in the pure flushing case is acidity constant and independent of discharge.

When acidity load is plotted against discharge (Figure 5), the mathematical role of the flushing factor becomes clear, because F is the exponent.

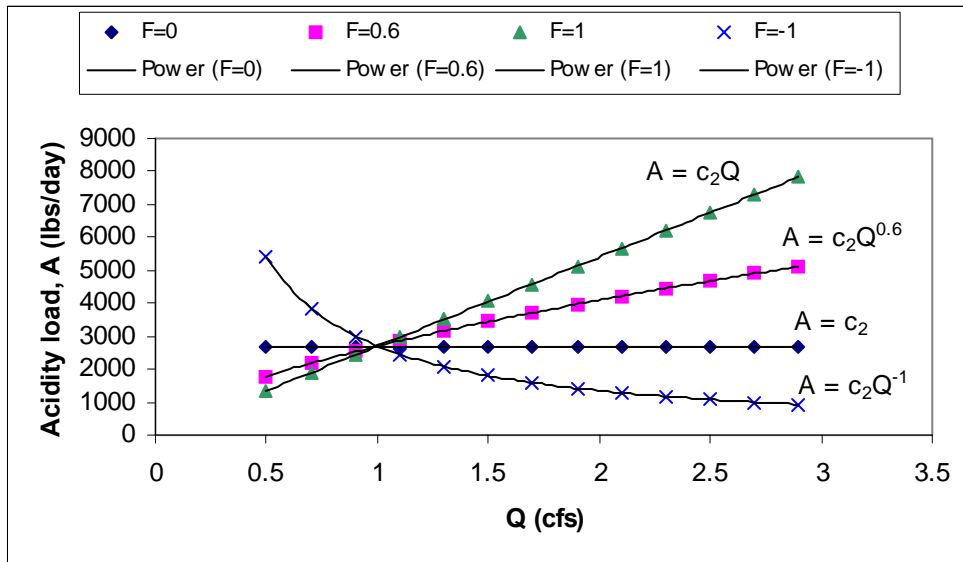


Figure 5. A plot of acidity load vs. discharge shows the role of the flushing factor (the exponent). When the behavior is sparing ($F < 0$), acidity load increases at a slower rate than discharge. When $F=0$, acidity load is constant and independent of discharge. When the behavior is flushing ($F > 0$), acidity load increases with discharge. $F=1$ is the special case where load increases in proportion to discharge, consistent with displacement of water from a uniform source such as a flooded mine. Partial flushing occurs when $0 < F < 1$.

By dividing all discharges by the mean annual discharge, they can be normalized, or made dimensionless. Assume a 1 cfs mean annual discharge for the hypothetical watershed. To linearize the power-law function, the appropriate transformation is into log-log space (Figure 6).

The relationship remains linear if acidity load ($A = a \cdot Q$), rather than concentration, is plotted on the y-axis (Figure 7).

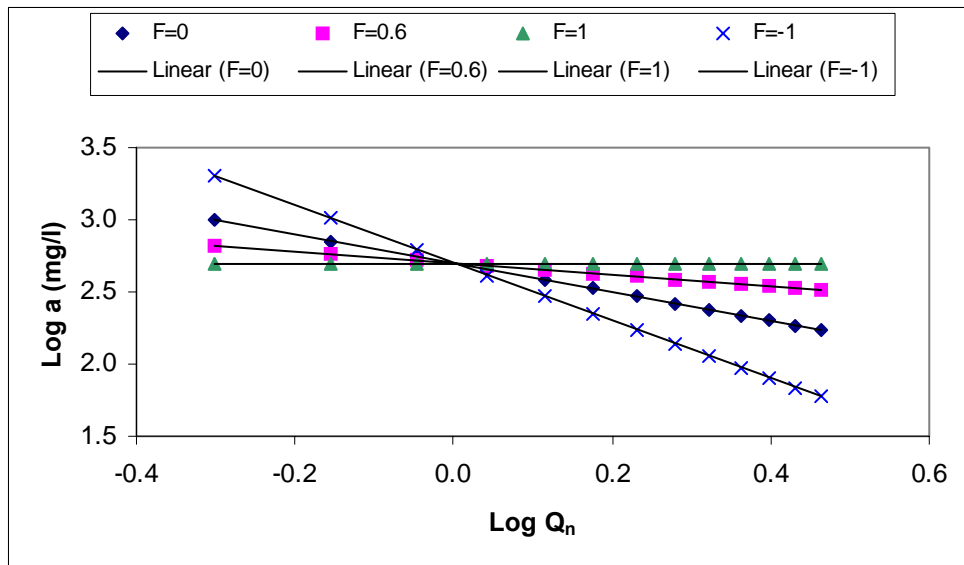


Figure 6. The acidity-discharge relationship is linearized by transforming both variables to log space, and the discharge is normalized by dividing it by mean annual discharge (assuming 1 cfs for the hypothetical watershed).

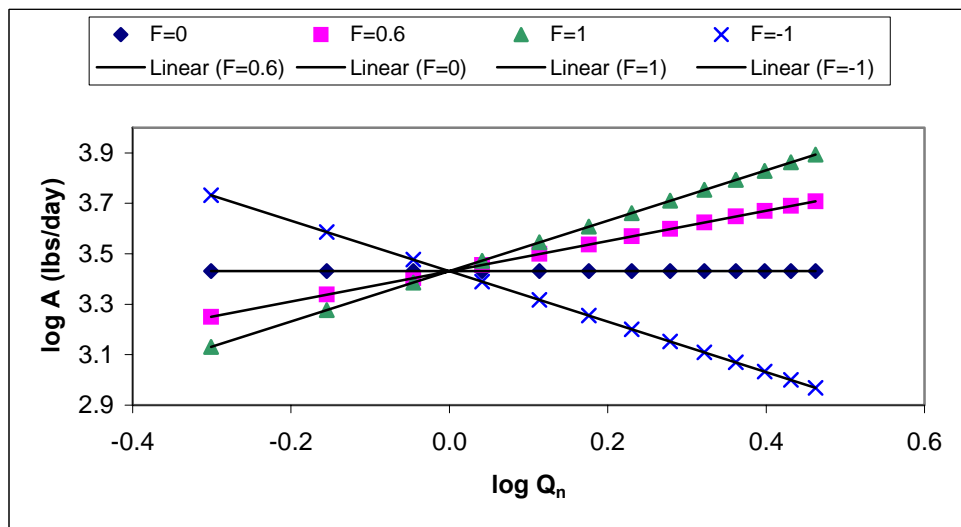


Figure 7. Multiplying acidity by discharge gives acidity load, plotted on the y-axis. The relationship remains linear in log space, for all cases. The slope indicates what mechanism controls acidity.

The hypothetical cases show general patterns that stream samples are expected to follow, depending on the predominance of storage vs. acidity production in a system. When stream samples are collected and analyzed before and after treatment, the pre-treatment data must be fitted to a function separately from the post-treatment data, because in general the processes will

be altered by reclamation. Comparing the two functions may show differences both in the magnitude of the source load and in the flushing factor (i.e., mechanism). To illustrate, the hypothetical partial-flushing case ($F=0.6$) is compared to a hypothetical post-treatment data set with pure flushing ($F=1$) and a decrease in the source load from 500 to 150 cfs·mg/l, or a decrease of 350 cfs·mg/l, which is 1890 lbs/day (Figure 8). Because the discharges are normalized, the mean annual discharge occurs at $\log Q_n = 0$. By using Excel to insert a linear trendline and reveal its equation, the intercepts are easily determined. The intercepts are the mean annual acidity loads before and after treatment, and the change ΔA , which is the performance measure, is the difference between the intercepts. Because they are in log space, it is necessary first to transform them: $\Delta A = 10^{3.43} - 10^{2.91} = (2691 - 813) \text{ lbs/day} = 1878 \text{ lbs/day}$. This example confirms that the method operates as expected, because this value is the same as the assumed change in source load, allowing for rounding error.

The log transformation introduces bias, because the scatter of data about the regression line is normally distributed in log space, but log-normally distributed in arithmetic space (Ferguson, 1986a). As a consequence, the transformation results in estimating the median rather than the mean value of acidity load for each discharge, which tends to underestimate the acidity load. A correction factor can be used, but other biases may remain (Koch and Smillie, 1986) due to violation of other assumptions (Ferguson, 1986b).

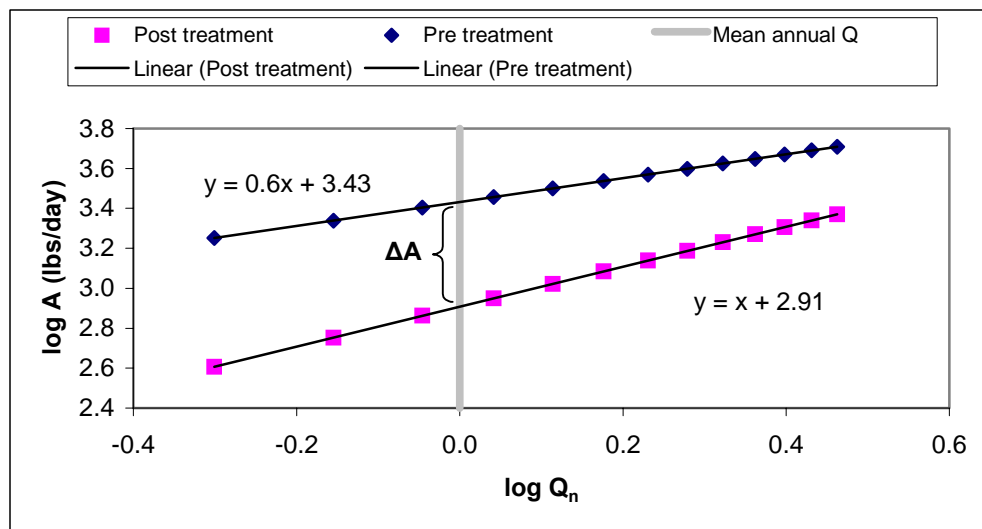


Figure 8. The hypothetical partial-flushing pre-treatment case ($F=0.6$) is compared to a hypothetical post-treatment pure-flushing case ($F=1$), where source load decreases from 500 cfs·mg/l to 150 cfs·mg/l. The “change in mean annual acidity load” ΔA is the difference in the y-intercepts after transformation, or $(10^{3.43} - 10^{2.91}) \text{ lbs/day} = 1879 \text{ lbs/day}$.

To normalize discharge, it is divided by mean annual discharge of the sampled stream. Mean annual discharge is a hypothetical discharge that represents a daily average of the total annual stream discharge for all years of record. In reality, this discharge is a moderate event that one seldom encounters. Instead, the stream is likely to flow at a lower discharge most of the time, but large events carry enormous volumes that skew the mean. It is a useful design concept because it represents how much water has to be treated in a given year.

Mean annual discharge (Q) for a site with known drainage area (A) can be estimated from USGS gauging data for streams in the Western Allegheny Plateau with long records, regressing Q against A (Figure 9). Watersheds are nested, with many smaller watersheds contained in a few larger ones, so the distribution of watershed area tends to be logarithmic, as seen by the clustering of points to the left.

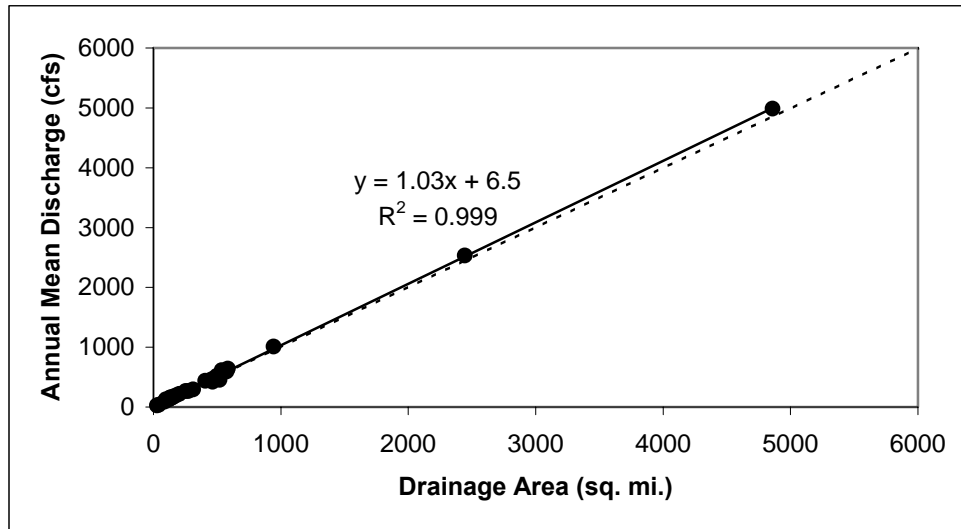


Figure 9. Mean annual discharge for southeastern Ohio basins is linearly related to drainage area in a ratio of $\sim 1\text{cfs}/1\text{ mi}^2$. Data points represent gage stations in the basins of the Muskingum River, Hocking River, Raccoon Creek, Beaver River, Little Beaver Creek, Yellow Creek, Short Creek, Shade River and Captina Creek (<http://nwis.waterdata.usgs.gov/nwis>). Only gages with record lengths exceeding 25 years are included.

Transformation to log-log space is appropriate, and spreads the data out so that the largest watershed does not unduly influence the regression (Figure 10). Regressing the transformed variables, and showing the equation (in Excel, for example), yields the parameters for predicting Q from A . Note that these parameters are in log space ($x = \log A$; $y = \log Q$), and must be transformed back to arithmetic space to estimate Q from A . By coincidence, streams in the Western Allegheny Plateau have a mean annual discharge of about 1cfs per square mile. While the regression is clearly significant, the variance of an individual estimate is of interest when

using the proposed method. To provide an idea of the magnitude of error, the median standard deviation for a discharge estimate using this regression is 28 cfs, or about 11% of the median measured discharge (267 cfs).

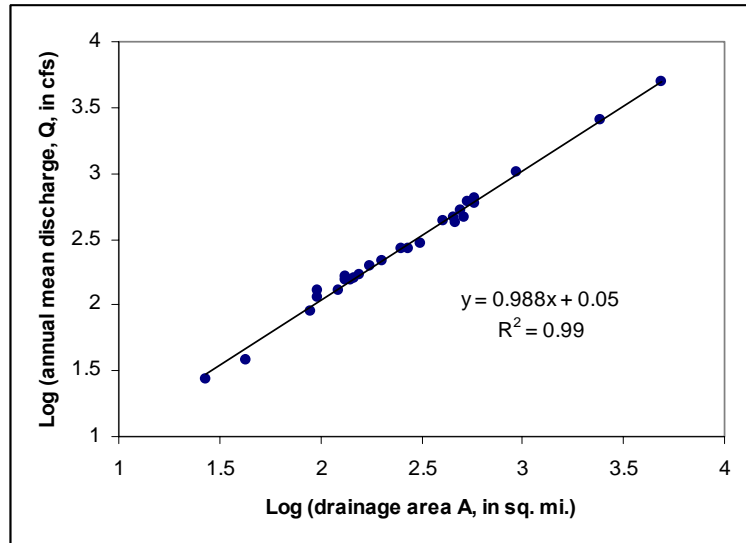


Figure 10. Data from previous figure are transformed to log space.

RESULTS AND DISCUSSION

The acid test of the proposed method is whether or not acidity in mined areas behaves according to a power-law model, and is thus amenable to the proposed analysis. In this section, the method is applied to streams of the Western Allegheny Plateau ecoregion that meet several criteria: stream waters contain large amounts of acidity due to coal mining; treatment for acid mine drainage has occurred, affecting acidity; and data density is sufficient to discern the shape of the acidity vs. discharge function.

For this paper, four projects are included that satisfy these criteria (Table 1).

Table 1. Projects analyzed in this study all were treated for AMD, using one or more methods as indicated. Cap = Refuse capping and revegetation; SAPS = successive alkalinity-producing system (wetland); Div. = stream diversion; OLC = open limestone channel; Dose = lime sand dosing; and Slag = steel slag leach beds. Start and end dates are for project construction. A = area of watershed contributing to the sample station.

Project	Cap	SAPS	Div.	OLC	Dose	Slag	Start	End	A (mi ²)
Buffer Run	x	x	x	x		x	Jul 1998	Sep 1999	1.99
Rock Run 24				x			2001	2001	0.205
Brush Fork					x		Oct 1997	Jun 2000	3.00
Flint Run	x	x	x	x		x	1984	1987	4.09

Buffer Run data (Figure 11) illustrate the method as applied to actual field data. At this site in Raccoon Creek in 1998 and 1999, refuse was capped and surface water was diverted from refuse, and several passive treatment systems were constructed, including a successive alkalinity-producing system (SAPS), an open limestone channel (OLC) and steel-slag leach beds (Davis, 2002). Acidity and discharge, when plotted vs. time (a), both show high variability. Although it is clear that acidity has decreased with time, a regression of the acidity data is weak. Moreover, acidity appears inversely correlated with discharge, and it is desirable to compensate for this dependence. Acidity is plotted as a function of discharge separately for pre- and post-treatment data subsets (b), and shows that the relationship follows a power-law model. When plotted as acidity load vs. discharge (c), the exponents become the flushing factors F . In this case, $0 < F < 1$, indicating a partial-flushing model. That is, load increases with discharge, but not as fast as discharge. Some dilution is occurring at higher flows. After treatment, the behavior moves toward a lower mean acidity load (decrease of over 2000 lbs/day) and more of a pure dilution model, consistent with less oxidation products flushing from unsaturated refuse under high-flow conditions. The decrease is $10^{3.40} - 10^{2.63} = 2100$ lbs/day.

Flint Run's primary contaminant source is a partially reclaimed and revegetated valley-fill coal waste pile capped by an organic papermill byproduct, Mead Bypro (Laverty, 2004). Analysis by the mean annual acidity loading method (Figure 12) shows that, since partial reclamation, the behavior has made a transition from almost pure flushing (loading increasing at the same rate as discharge) to a partial-flushing model, consistent with refuse capping. The load reduction is $10^{4.32} - 10^{3.55}$ lbs/day, or 17,000 lbs/day.

Rock Run 24 was treated by construction of an open limestone channel in 2001. Decreases in acidity are modest, from $10^{2.23}$ to $10^{2.05}$, or 170 to 112 lbs/day. The site is interesting because it shows unusual flushing behavior with $F > 1$: Acidity load *increases* at higher discharges. No source control was done at this site, so the flushing behavior is not surprising. The project is also interesting because it shows the possibility of detecting even small changes using the proposed method. Because the method explicitly accounts for the variability due to discharge, the primary source of variability, time variation is discernible.

Brush Fork was treated by limesand dosing (Figure 1) from 1997 to 2000. Frequent data collection during dosing allowed the data to be divided into three sets: pre-dosing, early post-dosing, and late post-dosing. Because the dosing was expected to add limestone storage to the

creek sediments, the acidity reduction was expected to increase with time. Analysis by the proposed method (Figure 14) successfully segregates the data into three series. The flushing factor remains constant as expected ($F = 0.77$) because dosing is not a source-control method. The loading changes from $10^{3.88}$ (that is, 7600) lbs/day to $10^{3.53}$ (or 3400) lbs/day to $10^{3.368}$ (or 2300) lbs/day.

The method can also be applied to mainstem sampling stations in watersheds undergoing restoration as well as natural attenuation. Application to Little Raccoon Creek (Figure 15) is illustrative. Data collected at the State Route 325 site, with a drainage area of 154 mi^2 , show a transition from an almost pure flushing model ($F=0.84$) to a partial flushing model ($F = 0.7$). The load reduction is 15,800 lbs/day (2900 tons/year).

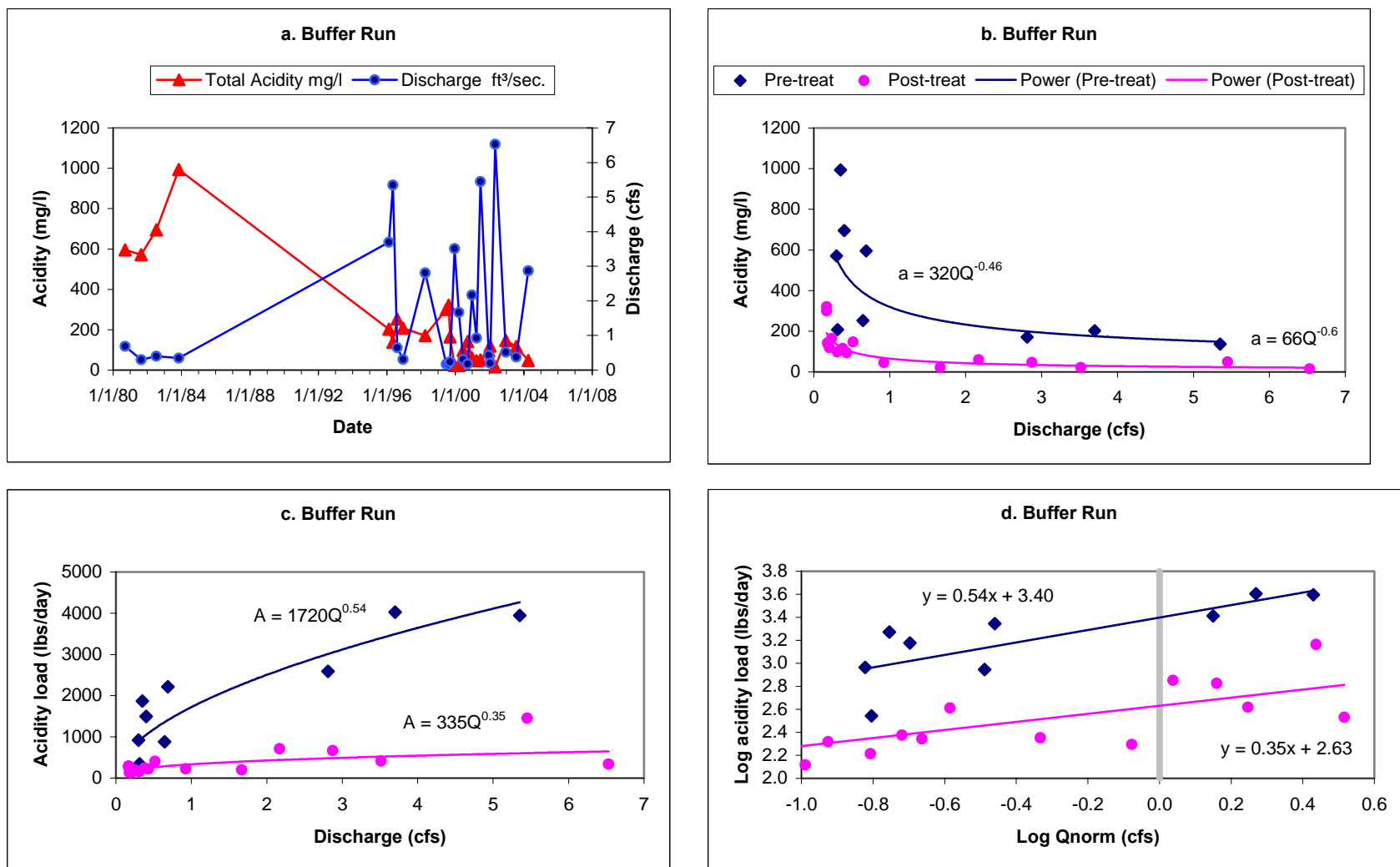


Figure 11. Buffer Run acidity and discharge data (a) show high variability. Segregating data into pre- and post-treatment subsets and plotting as *acidity* vs. discharge (b) shows the power-law behavior. Plotting as *load* vs. discharge (c) yields power-law equations with exponents F , in this case $0 < F < 1$, or partial –flushing behavior. In log-log space (d), the functions become linear and the decrease is found at mean annual flow (log $Q_{norm} = 0$).

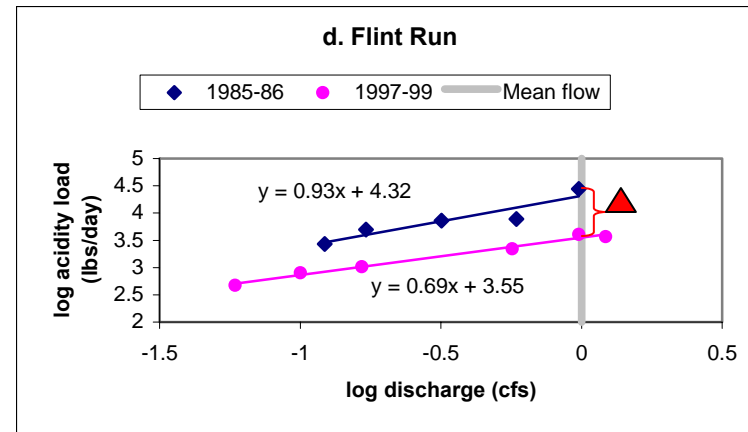
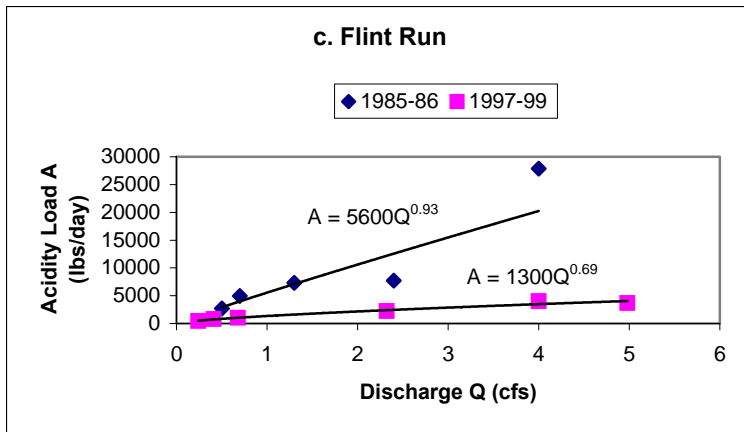
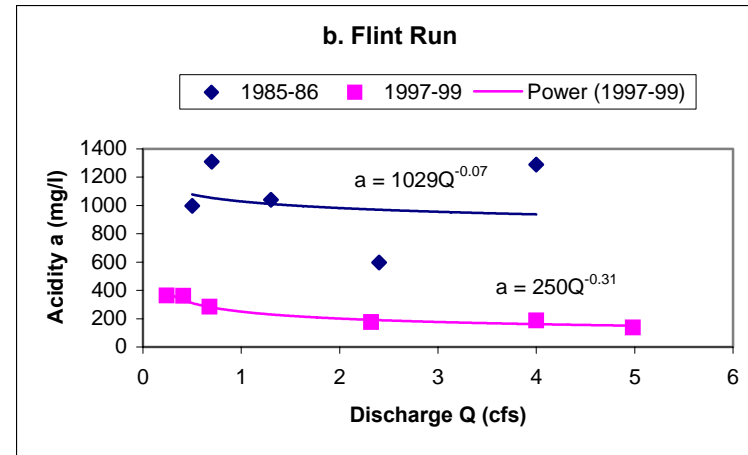
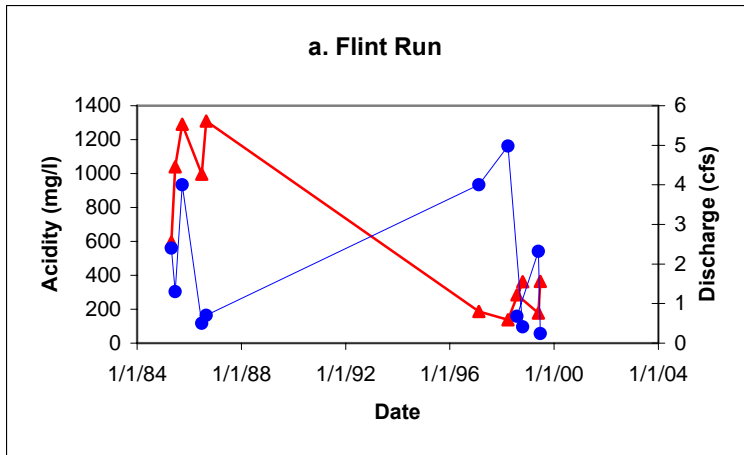


Figure 12. Flint Run, a refuse pile that was partially treated by capping with Mead Bypro in the mid-1980's, shows a large decrease in acidity (a). Behavior has made a change from almost pure flushing to partial flushing. The decrease in acidity load is 17,000 lbs/day (from $10^{4.32}$ to $10^{3.55}$; see triangle).

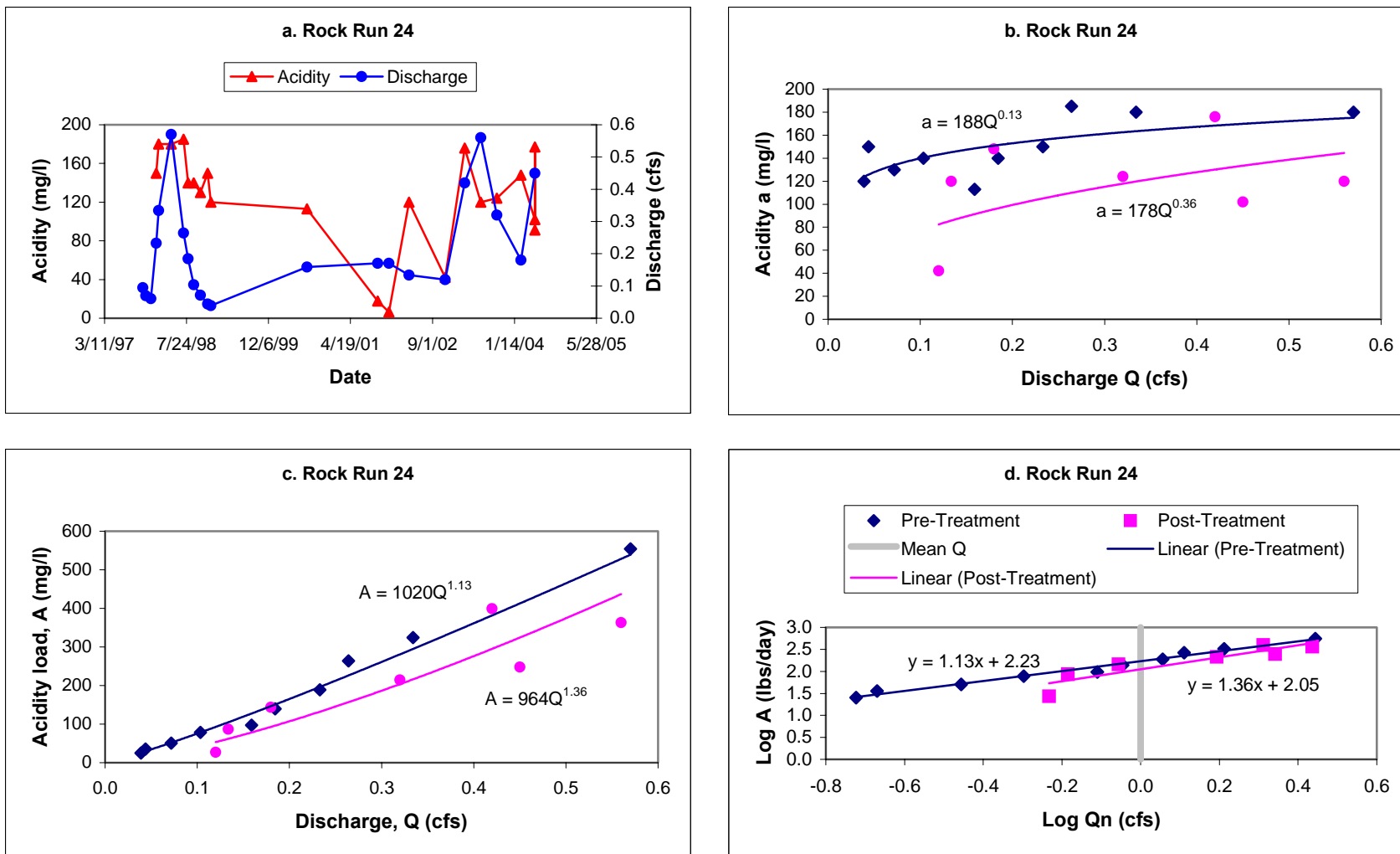


Figure 13. Rock Run 24 was treated with just an open limestone channel in 2001. Decreases in acidity are modest, from $10^{2.23}$ to $10^{2.05}$, 170 to 112 lbs/day. The site is interesting because it shows unusual flushing behavior with $F > 1$: Acidity load *increases* at higher discharges. No source control was done at this site, so the flushing behavior is not surprising. The project is also interesting because it shows the possibility of detecting even small changes.

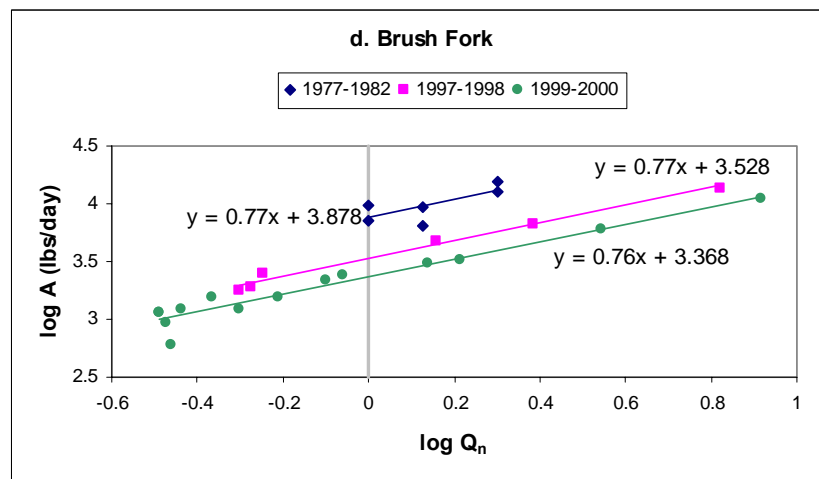
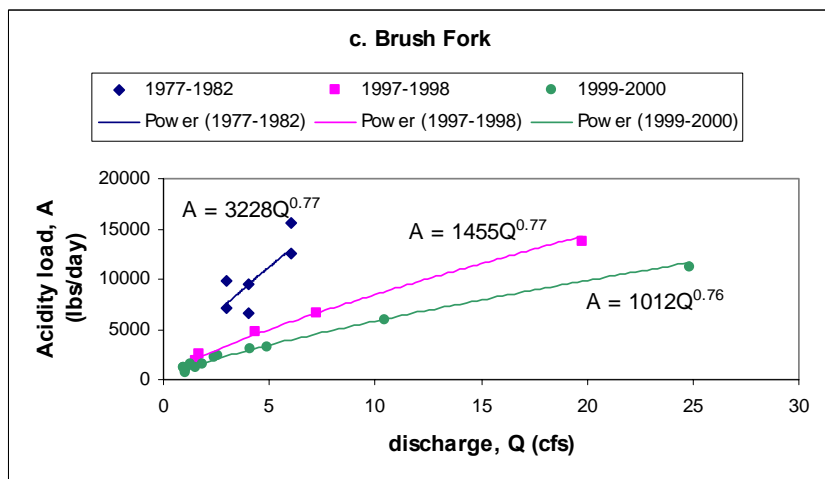
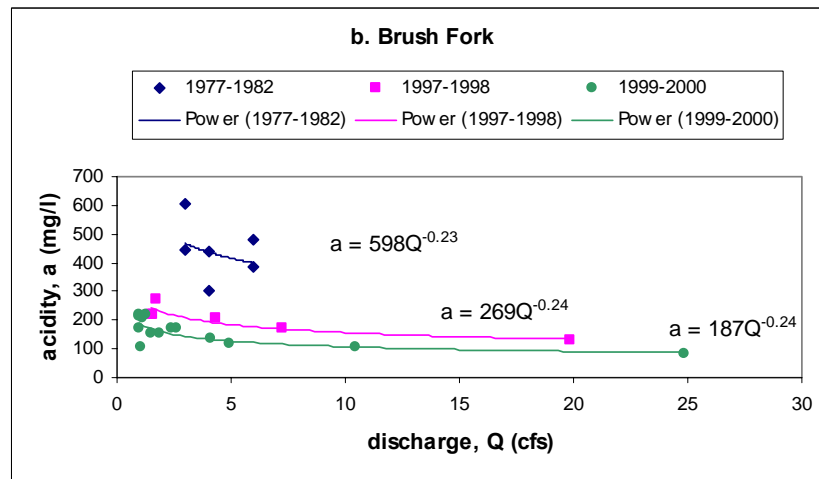
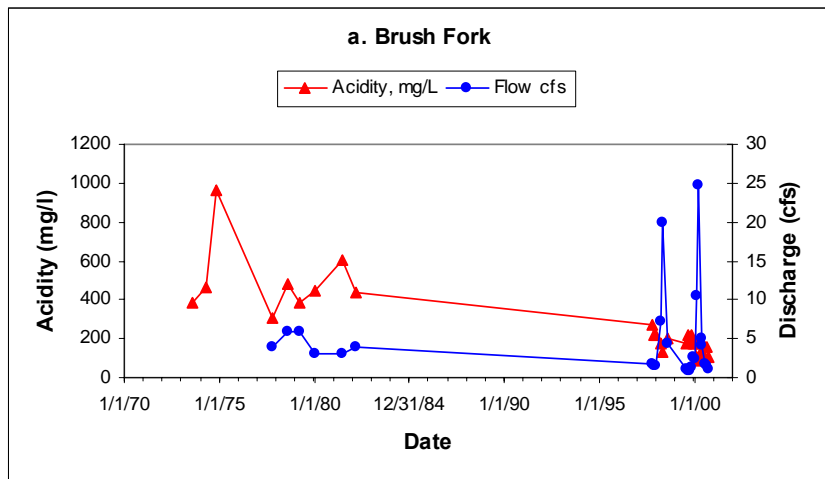


Figure 14. Brush Fork was treated by lime sand dosing from 1997 to 2000. This case illustrates the possibility of discriminating several time periods, pre-treatment, treatment, and late treatment, reflecting accumulation of limesand in the stream bed.

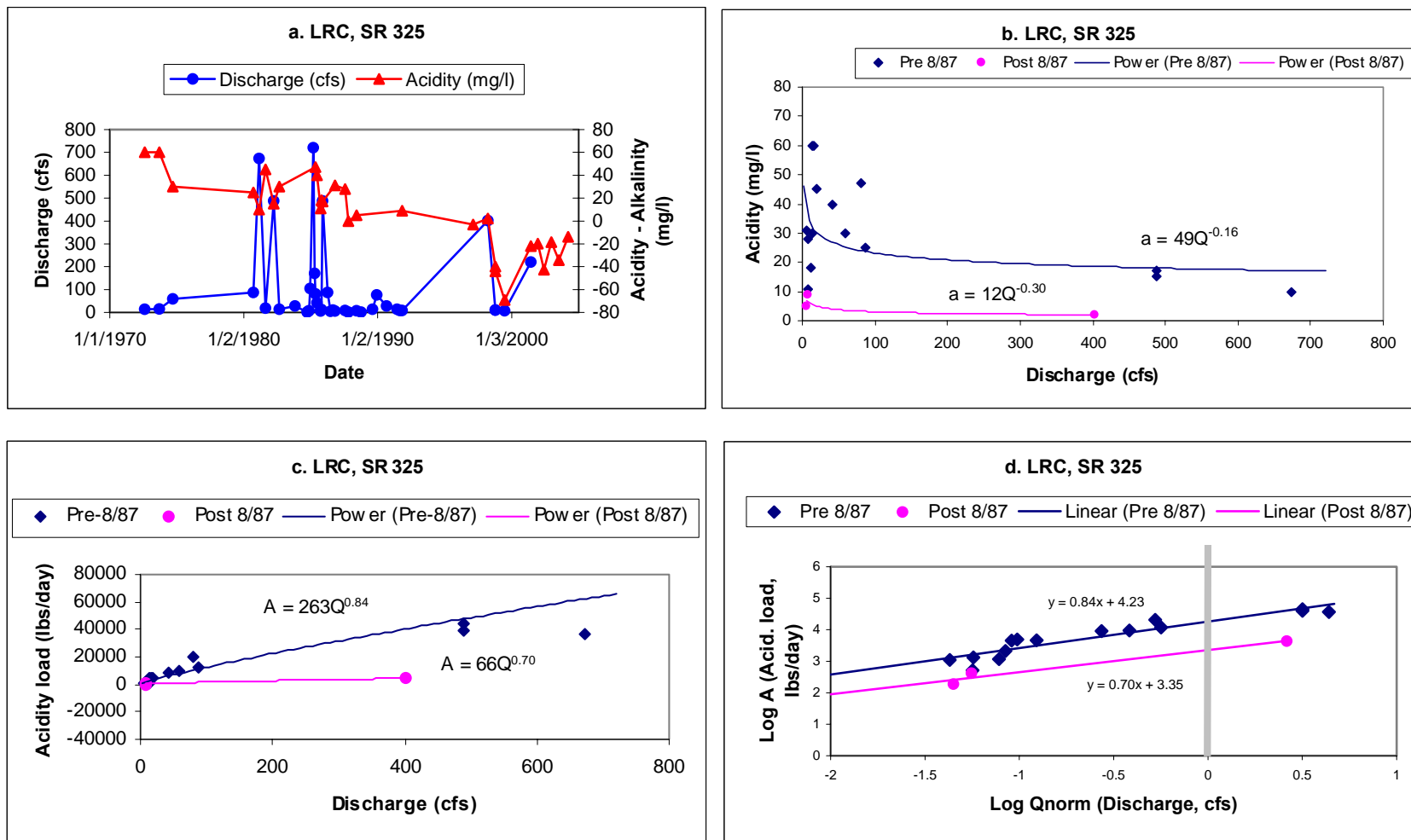


Figure 15. Acidity load decreases for a 154 mi² drainage area along the mainstem of Little Raccoon Creek demonstrate the applicability of the method to large areas with multiple projects as well as natural attenuation. The overall decrease in loading is 15,800 lbs/day ($10^{4.26} - 10^{3.35}$ lbs/day).

CONCLUSIONS

Acidity loads (A) in streams of southeast Ohio that are affected by AMD generally follow a power-law model, $A = c/Q^F$, where Q is discharge and exponent F is defined as a “flushing factor.” Flushing occurs ($F > 0$) when the contaminant load increases with discharge. Most commonly, partial flushing occurs ($0 < F < 1$): Load increases at a lesser rate than discharge. Less commonly, load increases *faster* than discharge ($F > 1$), perhaps due to washing of reaction products from mine walls or the unsaturated zone of a refuse pile. *Load* can be constant ($F = 0$) if there is a constant pollutant source, and this case is called a dilution model. *Concentration* also can be constant ($F = 1$) if load increases in proportion to flow. The constant-concentration case may occur if streamflow is derived from displacement of water from a uniformly polluted source such as a partially flooded mine, for example. Another possible behavior is “sparing” behavior ($F < 0$), in which load decreases as discharge increases. The mechanism could be extinguishing of acid-generating reactions at high flow, perhaps due to flooding of reactive sites.

Transformation of the power-law function of load vs. discharge into log-log space yields a linear function. This linear relationship yields a graphical tool for estimating acidity load at any discharge. For purposes of performance assessment, the mean annual discharge is the discharge of choice because it is the average of all probable daily discharges over a typical year. The mean annual load, which corresponds to the mean annual discharge, is similarly the average of all probable daily loads over a typical year. The mean annual load is clearly a proxy for the overall annual load from a stream or tributary. The linear relationship can be constructed for subsets of data from before and after treatment. Comparing the mean annual loads for each subset yields the performance measure: decrease in mean annual load.

Several caveats are in order. The method is valid only if the mean annual discharge is stationary in time. Climate change or major land use change can alter the watershed hydrology and change the mean annual streamflow. Mining can alter the natural water budget if underground mines transport water across watershed boundaries. Restoration can alter the water budget if subsidence closure restores surface drainage. Where such conditions are known, drainage areas (and therefore mean annual discharge) can be adjusted accordingly. If acidity load is highly dependent on discharge, that is, $|F|$ is large, then the change in acidity load may be overestimated or underestimated by assuming that the center of mass of the water is the center of mass of the acidity.

Reduction of acidity load is not a sufficient measure of stream restoration success, especially as loads decrease to near zero. It should be used together with biological criteria (OEPA, 1988). The mean annual acidity load can be used as an index to measure success in alleviating the dominant stressor, but other stressors may impair biological performance. Load-duration curves may be needed to interpret biological criteria, because they can explain duration of excursions from tolerable conditions. Isolation of stream reaches may be a cause for impairment even with no acidity (Stoertz et al., 2002). Excursions of key water-quality parameters from tolerable conditions during flushing events or droughts may limit biological communities (e.g., Archer and Newson, 2002). The method described here does not provide information on the duration or frequency of extreme events, but neither do many statistical approaches. Instead, they test hypotheses about the center of a continuous distribution (NRC, 2001, p. 62). Once the loading target has been achieved, if biotic indices (e.g., IBI and ICI) do not attain expected levels, then other factors including isolation, load duration, habitat quality, and other contaminants will have to be examined.

REFERENCES CITED

- Antonopoulos, V.Z.; Papamichail, D.M., and Mitsiou, K.A., 2001, Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece: *Hydrology and Earth System Sciences* Vol. 5(4), pp. 679-691.
- Archer, D. and M. Newson, 2002. The use of indices of flow variability in assessing the hydrological and instream habitat impacts of upland afforestation and drainage. *Journal of Hydrology* 268(1-4): 244.
- Bernard, T. and J. Young, 1997. *Ecology of Hope: Communities Collaborate for Sustainability*. New Society Publishers, 233 p.
- Bonta, J.V. and Cleland, B., 2003. Incorporating natural variability, uncertainty and risk into water quality evaluations using duration curves: *Journal of the American Water Resources Association* Vol. 39(6), pp. 1481-1496.
- Bonta, J.V. and Dick, W.A., 2003, Impact of coal surface mining and reclamation on surface water chemical concentrations and load rates in three Ohio watersheds: *Journal of the American Water Resources Association* Vol. 39(4), pp. 793-815.

- Dann, M.S., J.A. Lynch, and E.S. Corbett, 1986. Comparison of methods for estimating sulfate export from a forested watershed: *Journal of Environmental Quality* 15(2), p. 140-145.
- Darken, P.F., C.E. Zipper, G.I. Holtzman, and E.P. Smith, 2002. Serial correlation in water quality variables: Estimation and implications for trend analysis: *Water Resources Research* 38(7), pp. _____
- Davis, C., 2002. *The effect of the Buckeye Furnace Reclamation Project on water quality in Buffer Run, Jackson County, Ohio*: M.S. Thesis, Ohio University, Athens, Ohio, 84 p.
- Environmental Protection Agency (EPA), 2003,
- Ferguson, R.I., 1986a. River loads underestimated by rating curves: *Water Resources Research* 22(1), pp. 74-76.
- Ferguson, R., 1986b. Reply: *Water Resources Research* 22(13), pp. 2123-2124.
- Gibbons, R.D., 1994, *Statistical Methods for Groundwater Monitoring*, John Wiley & Sons, New York, pp. 175-185.
- Johnson, N.M., Likens, G.E., Bormann, F.H., Fisher, D.W., and Pierce, R.S., 1969. A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire: *Water Resources Research* Vol. 5(6), pp. 1353-1363.
- Karr, J.R., 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6(6): 21-27.
- Karr, J.R. and Yoder, C.O., 2004. Biological assessment and criteria improve Total Maximum Daily Load decision making: *Journal of Environmental Engineering* 130(6).
- Kendall, M.G., 1975. *Rank Correlation Methods* 4th ed., Charles Griffin, London.
- Koch, R.W. and G.M. Smillie, 1986. Comment on "River loads underestimated by rating curves" by R.I. Ferguson: *Water Resources Research* 22(13), pp. 2121-2122.
- Lacroix, G.L., 1987. Fish community structure in relation to acidity in three Nova Scotia rivers. *Canadian Journal of Zoology* 65(12): 2908.
- Laverty, B., 2004. *Hydrologic mechanisms in a coal refuse pile and their effects on acidity and metals loads to Little Raccoon Creek, Ohio*. M.S. Thesis, Ohio University, Athens, Ohio, 286 p.
- Lewis, W.M. and Grant, M.C., 1979. Relationships between stream discharge and yield of dissolved substances from a Colorado mountain watershed: *Soil Science* Vol. 128(6), pp. 353-363.

- Liebermann, T.D., Middelburg, R.F. and Irvine, S.A., 1987. *User's Manual for Estimation of Dissolved-Solids Concentrations and Loads in Surface Water*: U.S. Geological Survey Water-Resources Investigations Report 86-4124, __ pp.
- López, D.L. and Stoertz, M.W., 2001. Chemical and physical controls on waters discharged from abandoned underground coal mines: *Geochemistry: Exploration, Environment, Analysis* 1, pp. 51-60.
- Mann, H.B., 1945. Nonparametric tests against trend: *Econometrics* 13, pp. 245-259.
- Montgomery, D.C. and Johnson, L.A., 1976, *Forecasting and Time Series Analysis*, McGraw-Hill, New York.
- National Research Council (NRC), 2001. *Assessing the TMDL Approach to Water Quality Management*, National Academy Press, Washington, D.C., 109 pp.
- Ohio Environmental Protection Agency (OEPA), 1988. *Biological Criteria for the Protection of Aquatic Life: Volume I: The Role of Biological Data in Water Quality Assessment*, Columbus, OH, 44 pp.
- Ohio Environmental Protection Agency (OEPA), 1987. *Biological Criteria for the Protection of Aquatic Life: Volume II: Users Manual for Biological Field Assessment of Ohio Surface Waters*, Columbus, OH, 278 pp.
- Ormsbee, L., A. Elshorbagy, and E. Zechman, 2004. Methodology for pH Total Maximum Daily Loads: Application to Beech Creek Watershed. *Journal of Environmental Engineering* 130(2): 167-174.
- Pigati, E.M., 1997. *The hydrology and geochemistry of an abandoned underground coal mine: The Majestic Mine, Athens County, Ohio*. M.S. Thesis, Ohio University, 168 pp.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau: *Journal of the American Statistical Association* Vol. 63, pp. 1379-1389.
- Skousen, J.G. and Ziemkiewicz, P.F., 1996. *Acid Mine Drainage Control and Treatment*, 2d ed., West Virginia University and the National Mine Land Reclamation Center, Morgantown, WV, 362 pp.
- Smith, R.A.; Hirsch, R. M.; and Slack, J.R., 1982. A study of trends in total phosphorus measurements at NASQAN stations: *U.S. Geological Survey Water Supply Paper 2190*, U.S. Geological Survey, Alexandria, VA.

- Smith, E.E. and K.S. Shumate, 1971. Rate of pyrite oxidation and acid production rate in the field. Proceedings of Acid Mine Drainage Workshop, Ohio University, Athens, Ohio, USA, p. 11-18.
- Spiegel, M.R., 1968. *Mathematical Handbook of Formulas and Tables*, McGraw-Hill, NY, 271 pp.
- Stachler, P., 1997. *Chemical and hydrologic variability of acid mine drainage from abandoned Esco #40 underground coal mine, Ohio*. M.S. Thesis, Ohio University, 103 pp.
- Stednick, J.D. and L.C. Roig, 1989. Kalman filter calculation of sampling frequency when determining annual mean solute concentrations. *Water Resources Bulletin* 25(3), p. 675-683.
- Stoertz, M., H. Bourne, C. Knotts, and M. White, 2002. Effects of isolation and acid mine drainage on fish and macroinvertebrate communities of Monday Creek, Ohio. *Mine Water and the Environment* 21(2): 60-72.
- Stoertz, M. W., M.L. Hughes, N.S. Wanner, and M.E. Farley, 2001. Long-term water quality trends at a sealed, partially flooded underground mine. *Environmental and Engineering Geoscience* VII(1): 51-65.
- Weisstein, E.W., 2003. *CRC Concise Encyclopedia of Mathematics*, 2d. ed., Chapman and Hall, Boca Raton, FL, 3242 p.
- Ziemkiewicz, P.F., J.G. Skousen, and J. Simmons, 2003. Long-term performance of passive acid mine treatment systems. *Mine Water and the Environment* 22(3): 118-130.