Mine Water Environ (2014) 33:177–186
DOI 10.1007/s10230-014-0269-5

TECHNICAL ARTICLE

Acidity Loading Behavior in Coal-Mined Watersheds

Natalie A. Kruse · Mary W. Stoertz · Douglas H. Green · Jennifer R. Bowman · Dina L. Lopez

Received: 30 August 2013 / Accepted: 5 March 2014 / Published online: 19 March 2014

Abstract In this paper, we propose a method (the Stoertz Water Quality Evaluation Method) to estimate the reduction in acidity load at mean annual discharge (l) in acid mine drainage impaired streams. For a given stream, acidity loading (L), defined as the product of measured acidity concentration and discharge (Q) for a range of streamflow conditions, is proportional to l and the ratio (Q_n) of Q and the mean daily discharge (Q), according to the following equations: $L = l\ (Q/Q)^F = l\ Q_n{}^F$ and $\log L = \log\ l + F \log Q_n$, where F is a 'flushing factor' characterizing the stream loading behavior, such that $F = 0$ in pure dilution, $F > 0$ for flushing behavior, and $F < 0$ for sparing behavior. The assumed power-law relation between acidity load and discharge is supported by test cases. Acidity load at mean annual discharge can be a useful measure of the quantity of acid removed by mine drainage treatment, particularly where data for pre- and post-treatment cases can be compared. The proposed measure allows resource managers to assess and report the effectiveness of source control, passive treatment, and active treatment in alleviating acidity loads. This method can also be used to assess inputs of contamination to streams compared to pristine or background conditions.

Keywords Acid mine drainage · Dilution · Flushing · Load · Total maximum daily load · Reclamation · Nonpoint source

Dr. Mary Wilder Stoertz inspired and built the watershed community in Southeastern Ohio. As a faculty member in Ohio University's Department of Geological Sciences and an active community member, she was the driving force behind the still thriving partnerships between communities, government agencies, and Ohio University. Mary passed away Feb. 26, 2007, but her work lives on in the hearts and minds of watershed communities in the Ohio minelands. This manuscript was being prepared at the time of her death.

N. A. Kruse (✉) · J. R. Bowman
Voinovich School of Leadership and Public Affairs, Ohio University, Athens, OH 45701, USA
e-mail: krusen@ohio.edu

M. W. Stoertz · D. H. Green · D. L. Lopez
Department of Geological Sciences, Ohio University, Athens, OH 45701, USA

Introduction

The US Congress enacted the Federal Water Pollution Control Act in 1972, commonly known as the Clean Water Act (CWA) after its 1977 amendment. The CWA set the regulatory structure for the discharge of pollutants into US waters under the authority of the US Environmental Protection Agency (USEPA). The CWA also stated the need for planning to address nonpoint source pollution (USEPA 2013a). Within the framework of the CWA, States specify the uses for each intrastate water body and the States or USEPA regulate discharges to ensure that these designated uses can be met.

States are required to maintain a list of water bodies that fail to meet water quality criteria for the designated uses, and are required to develop plans for alleviating contaminant loading to these water bodies. The total maximum daily load (TMDL) program is the primary planning tool for restoring and maintaining surface water quality in the USA (USEPA 2012). A TMDL is a loading at which, when divided by flow rate, the concentration of the constituent of concern meets the in-stream criteria. In other words, the TMDL is a calculation of the maximum amount of a pollutant that a water body can receive and still meet water quality standards after accounting for the flow rate at the monitoring location. A TMDL is the sum of the allowable loads of a single pollutant from all contributing point *and* nonpoint sources upstream of the monitoring location, including a safety margin and allowing for seasonal variability (USEPA 2012). While point-source loads can be measured relatively easily, nonpoint-source contaminant loads are more difficult to either measure or remediate.

The USEPA requires and the National Research Council (NRC 2001) recommends that States conduct a periodic assessment of their water bodies to determine if designated uses have been attained and maintained. On that basis, the USEPA (2013b) and States publish periodic water quality assessments that summarize the current status of attainment by water body, including relevant TMDLs for those that are designated as impaired. Limited budgets prevent States from routinely monitoring all indicators, even on major rivers, so a method that effectively uses scarce data for reasonable estimates of loads and load reductions is valuable (NRC 2001). Measuring progress in nonpoint-source load reduction has practical value: public managers (e.g. state agencies and local watershed groups) can plan restoration strategies to reduce remaining contaminant loads; state regulatory authorities and watershed managers can (and must) report progress to USEPA and funding sources; and managers can compare the efficiencies of restoration strategies.

Despite efforts by regulatory authorities, industry, and the public to reduce contaminant loads, 55 % of assessed US waters currently do not meet State water-quality standards (USEPA 2013b). 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. decreases in in-stream pollutant concentrations), and response indicators (e.g. biometrics) (Karr and Yoder 2004). Of these, only biometrics directly measure the end outcome; they have been referred to as the "gold standard endpoint" (Karr and Yoder 2004). Chemical criteria are not reliable indicators of biological condition, and so are not substitutes for biometrics. In a study by Davis et al. (1996), 25 % of river distances assessed in the United States were found degraded based on chemical criteria, but twice as much 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 versus passive treatment systems for acid mine drainage (AMD) 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 remediation activities (e.g. liming, fencing cattle, or eliminating a point source).

In AMD-impaired streams, the primary stressor is water quality. In some cases, physical aquatic habitat may be excellent apart from water chemistry (Stoertz et al. 2002). In other cases, the physical habitat may be degraded because of encrustation and embedded substrates resulting from the precipitation of iron and other contaminants (DeNicola and Stapleton 2002; Earle and Callaghan 1998). Acidity derives from the main characteristics of AMD, namely low pH and elevated concentrations of iron, manganese, and aluminum. These constituents are known to limit fish and macroinvertebrate communities (e.g. Allard and Moreau 1987; Kowalik et al. 2007; Kruse et al. 2012; Lacroix 1987; Rosemond et al. 1992). Acidity is an appropriate variable to assess partial recovery of acidic mining-impaired streams and support of their 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), impacts aquatic life, and is easy to measure by titration or by estimation from pH plus dissolved metals (Kirby and Cravotta 2005a, b; Skousen and Ziemkiewicz 1996). 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 do not require a normal distribution and allow missing data or unevenly spaced measurements and, thus, are commonly used for hydrological data analysis (Helsel and Hirsch 2002). Examples of non-parametric trend estimators are Sen's test (Sen 1968), the Kendall test (Kendall 1975; Mann 1945), and the seasonal Kendall test (Smith et al. 1982) that allows for seasonally variable data (Gibbons 1994).

A data set collected at a stream sampling station (Fig. 1) shows acidity over a 13 years period, with two dozen samples during that time, which on its own is insufficient to evaluate water quality improvement considering seasonality and changing hydrological conditions. Installation of a lime doser was completed in March 2004. Three samples were collected prior to treatment and 21 after treatment. A regression of the time series of acidity data versus time yields an r^2 of only 0.23 (Fig. 1). Acidity appears to have a relation with discharge: high acidities are measured during low discharge, and vice versa. High flows are often critical because of their associated high loads. However, monitoring is seldom conducted for nonpoint pollution sources during critical high-flow events (NRC 2001) and

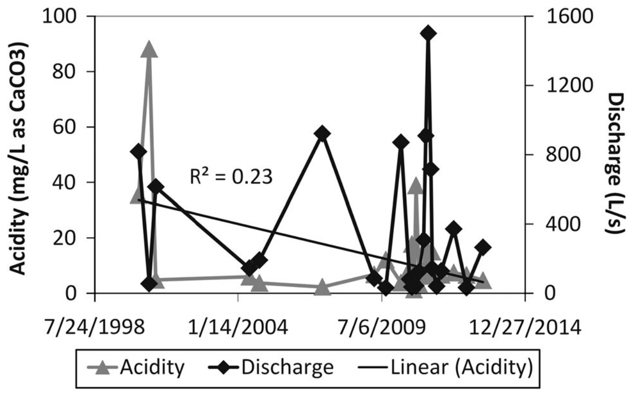

Fig. 1 This data set shows a trend in total acidity from 1999 (pre-treatment) to 2012 (post-2004-treatment). These data are variable and flow-dependent, however, so that a linear regression is misleading

consequently, a quantitative relation between discharge and acidity concentration is difficult to establish.

In this paper, we propose a way to estimate the reduction in acidity load at mean annual discharge in AMD impaired streams. By finding the difference between the acidity load at mean annual discharge for pre- and post-treatment conditions, this parameter becomes a performance measure that integrates probable acidity load changes over an entire year. We call this the "Stoertz Water Quality Evaluation Method"; it can be used to assess decreases in metal pollution in streams affected by neutral or basic mine drainage, as described in the Method Limitations and Further Applications section.

Relationship Between Acidity Concentration and Discharge

Several patterns or models relating solute concentrations and discharges exist. The different patterns result because stream water is a mixture of contributions from various sources that have different compositions and exhibit spatial and temporal variability (e.g. Cravotta et al. 2013; Johnson et al. 1969). Correlations between concentration and discharge can be produced by dilution as well other processes (e.g. evaporation, water exchange with groundwater). Concentrations may decrease (become diluted) with increased discharge if the main source of the constituent has a relatively constant composition, such as a point source or natural aquifer baseflow, and the streamflow increases in response to runoff or other inflows that are not sources of the constituent of interest. Concentrations at high discharges may also be 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." Sparing behavior can be interpreted as shutting down of reactions at high flow, perhaps due to flooding of reactive sites and lack of access to oxygen. Purging or flushing can increase concentrations with discharge because of storm "washoff" or the dissolution of accumulated oxidation products (e.g. Hirsch et al. 1982). Thus, loads can increase at the same rate as discharge, increase more slowly than discharge by a "sparing" mechanism, or increase faster than discharge, by a "purging" mechanism.

The rate of acidity production depends on availability of oxygen, the presence of neutralizing alkalinity, and the storage of water and reaction products. If the stream is formed by outflow from an underground mine or waste rock pile, concentrations can vary during the year within an order of magnitude while the discharge can increase during flushing events over several orders of magnitude (e.g. Pigati and López 1999). In the main stem of larger streams or rivers receiving AMD streams, dilution by runoff during storm events can produce a decrease in concentration. Deep mine pools may act as reservoirs, with relatively uniform concentrations independent of discharge because the water flows from a well-mixed source or it may have a seasonally fluctuating water level that washes pyrite oxidation products from the mine walls to be released slowly as baseflow (Liebermann et al. 1987; López and Stoertz 2001; Stoertz et al. 2001). In stream headwaters where deep mine discharge dominates the flow, concentrations may be almost independent of discharge. On the other hand, mine discharge may constitute a point source or baseflow source that is diluted by surface water. In Appalachia, mine discharges show strong seasonality in flow, but the flow is controlled by evapo-transpirative demand on an annual cycle rather than by precipitation (Pigati and López 1999; Stoertz et al. 2001, 2002). Mining or reclamation, which change these features, can alter the relation between discharge and concentration and even the sign of the regression (Bonta and Dick 2003).

For the data presented in this study, standard operating procedures for acidity and alkalinity analyses (SM2310B, SM2320B) were followed in accordance with USEPA and the Standard Methods for the Examination of Water and Wastewater (Clesceri et al. 1998). However, one can calculate net acidity using various methods (i.e. measured hot acidity, measured and calculated acidity) (Kirby and Cravotta 2005a, b; Hedin 2004). For example, as summarized by Kirby and Cravotta (2005b), Hedin et al. (1994) and Watzlaf et al. (2004) describe calculated acidity in units of mg/L as $CaCO_3$ (Eq. 1):

$$\mathrm{Acidity_{calc}} = 50\left[\left(2\mathrm{Fe}^{2+}/56\right) + \left(3\mathrm{Fe}^{3+}/56\right) + \left(3\mathrm{Al}/27\right) + 2\mathrm{Mn}/55 + 1000\left(10^{-\mathrm{pH}}\right)\right] \quad (1)$$

where metal concentrations are in mg/L in Eq. (1).

Acidity loading may be calculated by multiplying acidity and discharge and converting into standard units (e.g. kg/day). Equation 2 illustrates this calculation for metric units

$$Q\ (\mathrm{L/s}) \times\ \mathrm{Acidity}\ (\mathrm{mg/L}) \times 0.0864 = \mathrm{Acidity\ loading}\ (\mathrm{kg/day}) \quad (2)$$

Mean Annual Discharge Determination

The mean daily discharge or mean annual discharge (Q_{mean}) is a useful design concept because it represents how much water exits the watershed or has to be treated in a given day or year. In reality, the mean annual discharge is a moderate estimate of streamflow that is weighted by infrequent high-flow conditions. The stream will flow at a lower rate most of the time, but large events carry enormous volumes that skew the mean. Measuring the mean value of discharge over a year requires a continuous record of streamflow. Although continuous record data may be available for stream gages operated by resource agencies such as the US Geological Survey (USGS) (http://waterdata.usgs.gov/nwis/sw), utilities, or others, such stream gages may not be located in the immediate watershed of interest.

To estimate mean annual discharge (Q_{mean}) at ungaged sites, USGS provides the online Streamstats program (Ries et al. 2008, http://streamstats.usgs.gov/). Streamstats provides convenient and easy access to obtain drainage area for user-specified basins across the USA as well as values of estimated mean annual discharge. For example, rural, forested streams in the Western Allegheny Plateau have a streamflow yield of about 10.9 L/s per km^2 or 1 cubic foot per second per square mile (Koltun and Whitehead 2002). At sites where there is no continuous long-term record of discharge at the sampling point, the mean daily discharge may be estimated by multiplying the drainage area of the sampling point by the average annual yield (discharge/area) for a continuous streamflow recorded nearby. The mean annual discharge (Q_{mean}) for a site with known drainage area (A) can then be estimated from gaging data for streams with long records by regressing Q_{mean} against A to develop an average annual yield for an area. The applicable discharge record site(s) generally should have comparable watershed characteristics (area, relief, land cover, geology, climate) as the ungaged site.

Acidity Load at Mean Discharge

The proposed 'Stoertz Water Quality Evaluation Method' is used to calculate the acidity load at mean annual discharge (l) and, theoretically at least, incorporates all likely acid loadings over a water year. Positive non-zero values of acidity loading and discharge divided by mean annual discharge are both log transformed before using a linear regression to determine the estimated acidity loading at mean annual discharge or the 'mean annual load.' The Stoertz Method uses the mean annual load as a tool for practitioners to compare loadings before and after an intervention or to determine the baseline state of a water body. Since the mean annual load is a proxy for the average of all daily loads, it provides a consistent measure of water quality accounting for discharge variation over a water year.

For a given stream, acidity loading (L), defined as the product of concentration and discharge (as described

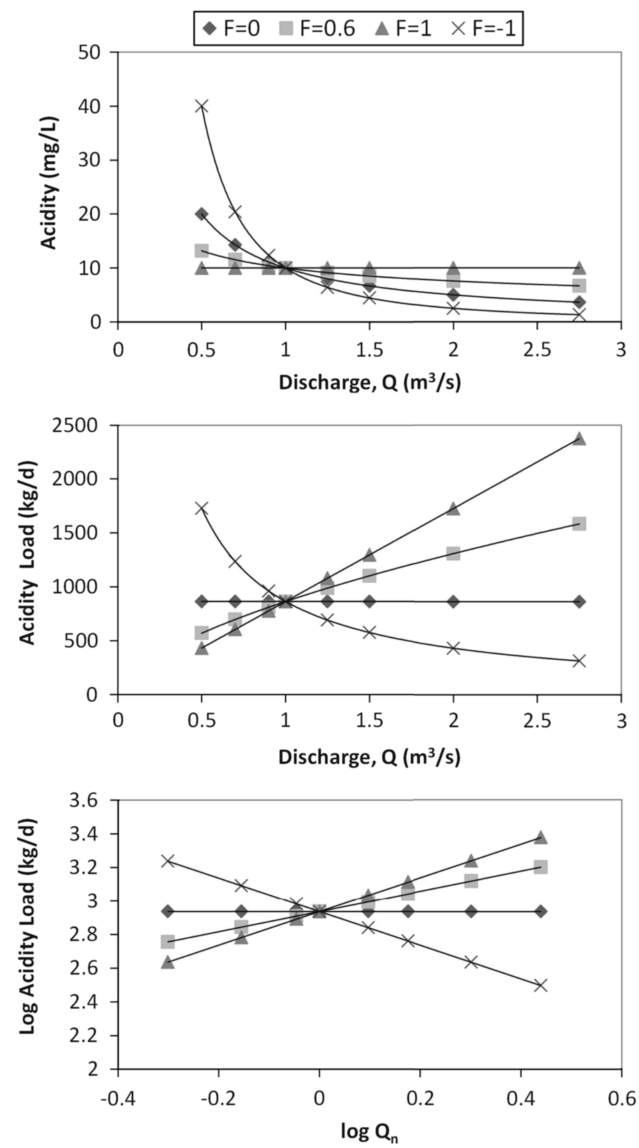

Fig. 2 **a** Hypothetical cases of pure dilution ($F = 0$), pure flushing ($F = 1$), partial flushing ($0 < F < 1$), and sparing ($F < 0$). **b** A plot of acidity load versus discharge shows the role of the flushing factor (the exponent). **c** Log normalized acidity plotted against log normalized Q_n (discharge divided by mean annual discharge) linearizes the relationship; the slope indicates what mechanism controls acidity. Note that negative values of acidity cannot be log transformed

above), is assumed in the Stoertz Method to depend on mean daily discharge $\mathcal{Q}$, according to:

$$\mathrm{L} = l(\mathrm{Q}/\mathcal{Q})^{\mathrm{F}} = l\mathrm{Q}_{\mathrm{n}}^{\mathrm{F}} \tag{3}$$

$$\log \mathrm{L} = \log l + \mathrm{F} \log \mathrm{Q}_{\mathrm{n}} \tag{4}$$

where $\mathcal{Q}$ is the mean daily discharge, Q is measured discharge, Q_n is normalized discharge ($Q/\mathcal{Q}$), l is the acidity load at mean annual discharge and F is a flushing factor that characterizes the stream loading behavior (Stoertz and Green 2007). At a mean discharge $Q_n = 1$, the acidity loading L is given simply by l. The flushing factor F is the slope of the curve of log L versus log Q_n. $F > 0$ for flushing, $F = 0$ for pure dilution, and $F < 0$ for sparing behavior.

Dilution of a source of acidity with a constant concentration follows a power law (Fig. 2), e.g. doubling discharge (Q) halves the acidity concentration (load remains constant). Flushing occurs ($F > 0$) when the contaminant load increases with discharge. Theoretically, the concentration can be almost constant (load increases proportional to flow; $F \approx 1$), if, for example, the streamflow is derived from displacement of water from a uniformly polluted source, such as a partially flooded mine. Most commonly, partial flushing occurs: load increases at a lesser rate than discharge ($0 < F < 1$). Occasionally, loads can increase faster than discharge ($F > 1$), when reaction products are washed from mine walls or the unsaturated zone of a mine waste pile. Sparing behavior can be interpreted as shutting down of reactions at high flow, perhaps due to flooding of reactive sites and lack of access to oxygen.

The proposed parameter, *acidity load at mean annual discharge* (l), assumes that the acidity versus discharge relationship follows a certain physical model such that the log-transformed variables are linearly related and are amenable to linear-regression methods. Due to the nature of log transformation, this method is only amenable to positive values; the Method Limitations and Further Applications section details this limitation and ways to overcome it.

In this section, hypothetical cases are used to show possible relationships and to interpret them physically and mathematically. The method is then applied to streams in the Western Allegheny Plateau to verify that real-world examples are consistent with the proposed model. Dilution, flushing (purging), and sparing, along with partial flushing (which is commonly observed in the field), are all illustrated for a hypothetical case (Fig. 2a). When acidity load is plotted against discharge (Fig. 2b), the mathematical role of the flushing factor becomes clear, because F is the exponent. By dividing all discharges by the mean annual discharge of the sampled stream, they can be normalized, or made dimensionless. To linearize the power-law function, the appropriate transformation is into log–log space (Fig. 2c).

The hypothetical cases (Fig. 2) show general patterns that stream samples are expected to follow, depending on the predominance of storage versus 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 the processes will generally be altered by reclamation. Comparing the two functions may show differences in both the magnitude of the source load and the flushing factor (i.e. mechanism). Note that the slope of the curve log L versus log Q_n is F. 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 of 1,890 kg/day (Fig. 3). Because the discharges are normalized, the mean annual discharge occurs at log $Q_n = 0$. The intercepts can be easily determined by fitting a linear trendline to the data. The intercepts are the mean annual acidity loads before and after treatment, and the change, ΔL, which is the performance measure, is the difference between the intercepts. Because the data are in log (base 10) space, it is necessary first to transform them: $\Delta L = 10^{3.43} – 10^{2.91} = (2{,}691–813)$ kg/day = 1,878 kg/day. This confirms that the method operates as expected, because this value is the same as the assumed change in source load, allowing for rounding error. This method is similar in nature to double mass plots developed by the USGS (Searcy and Hardison 1960); however, the pre- and post-treatment data is handled in separate series to find differences between the datasets, and distinct pairs of log transformed Q_n and L are analyzed rather than cumulative values. Where a clear break between pre- and post-treatment data is not obvious, methods detailed by Searcy and Hardison (1960) could be used to find a statistically significant break point.

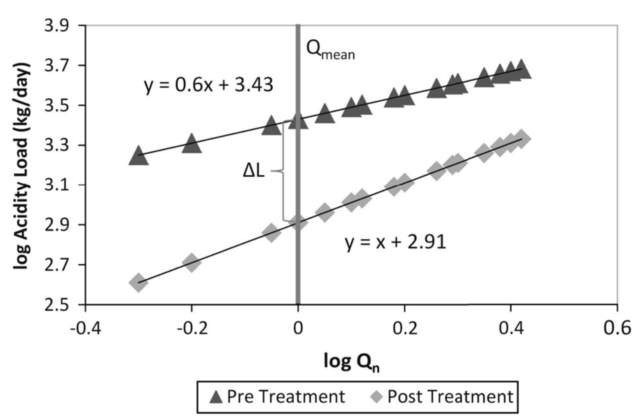

Fig. 3 The hypothetical partial-flushing pre-treatment case ($F = 0.6$) is compared to a hypothetical post-treatment pure-flushing case ($F = 1$). The "change in mean annual acidity load" ΔL is the difference in the y-intercepts after transformation, or ($10^{3.43}–10^{2.91}$) kg/day = 1,879 kg/day

Application to Case Studies

The acid test of the proposed method is whether or not acidity loading 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 eco-region that meet several criteria: stream waters contain large amounts of acidity due to coal mining; treatment for AMD has occurred, affecting acidity; and data density is sufficient to discern the shape of the acidity versus discharge function. For this paper, two projects are included that satisfy these criteria (Table 1).

The method was applied to actual field data from Jobs Hollow (Figs. 4, 5). A lime doser was installed at this site in Monday Creek (Perry County, Ohio) in 2004 to decrease the acidity load from the headwaters of Monday Creek by 54 % (Bowman 2011). When plotted versus time (Fig. 4), both acidity and discharge show high temporal variability. Although it is clear that acidity has decreased over time, a regression of the acidity data is weak. Acidity is plotted as a function of discharge separately for pre- and post-treatment data subsets (Fig. 5a), and shows that the relationship follows a power-law model. When plotted as acidity load versus discharge (Fig. 5b), the exponents become the flushing factors, F. In the pre-treatment case, $0 < F < 1$, indicating a partial-flushing model; load increases with discharge, but not as quickly, since some dilution is occurring at higher flows. After treatment, the behavior moves toward a lower mean acidity load and more of a pure flushing model, consistent with a point source (underground mine discharge) being treated at a constant rate. The linear regressions shown in Fig. 5c are statistically significant (pre-treatment $p = 1.8 \times 10^{-10}$, post-treatment $p = 2.4 \times 10^{-5}$). The decrease is $10^{2.50}$–$10^{1.45} = 288$ kg/day.

Table 1 Projects analyzed in this study were treated for AMD, using one or more methods as indicated

Project	Project type	Construction completion	Drainage area (km^2)
Jobs Hollow	Lime doser	2004	6.7
East Branch	Land reclamation, slag beds	2007	9.0

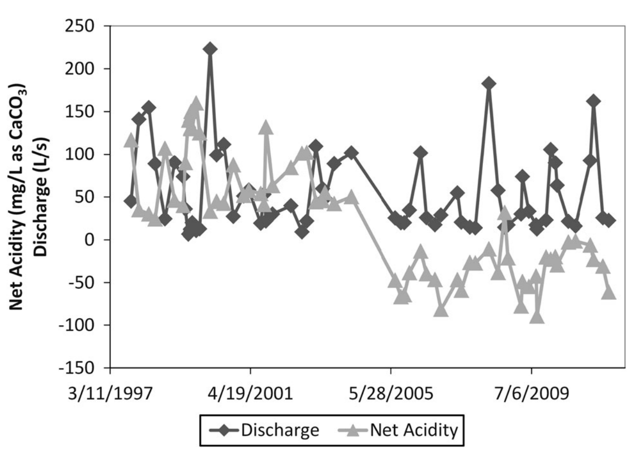

Fig. 4 Discharge (L/s) and net acidity (acidity–alkalinity) (mg/L as $CaCO_3$) over the period of record from 1997 to 2011 for the monitoring site downstream of the Jobs Hollow Doser

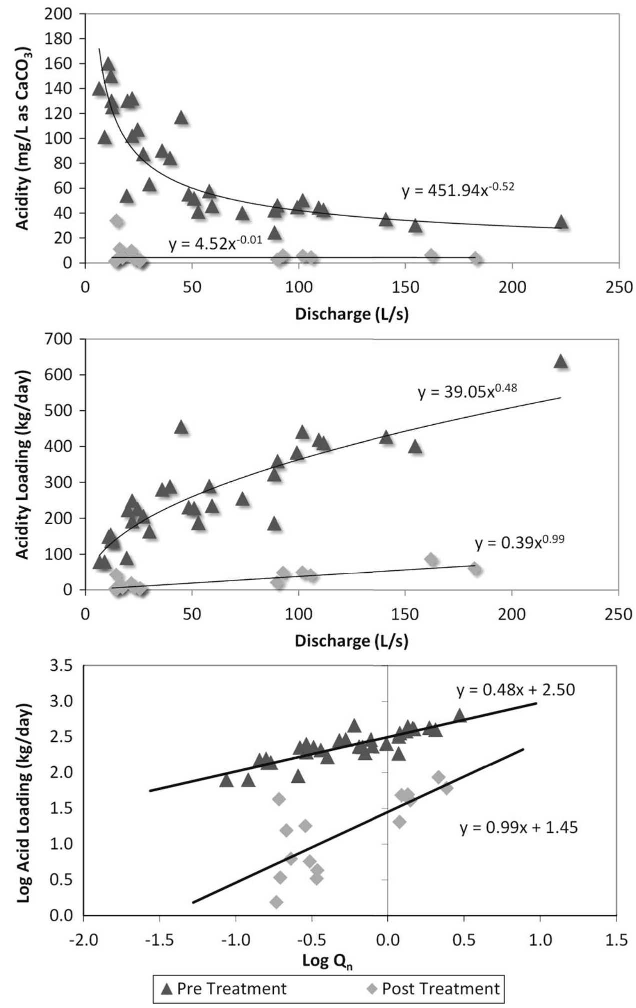

Fig. 5 **a** Total acidity (mg/L as $CaCO_3$) versus discharge (L/s) for pre and post-treatment data sets with power law regressions. **b** Acidity loading (kg/day) versus discharge (L/s) for pre and post-treatment data sets with power law regressions. **c** Log transformed acidity versus log transformed normalized discharge (discharge divided by mean annual discharge) with linear regressions

East Branch is a tributary to the headwaters of Raccoon Creek in Hocking County, Ohio and has been treated in two small sub-watersheds. The first phase project, which is evaluated here, consisted of strip mine reclamation and the installation of six steel slag leach beds, complete in 2007

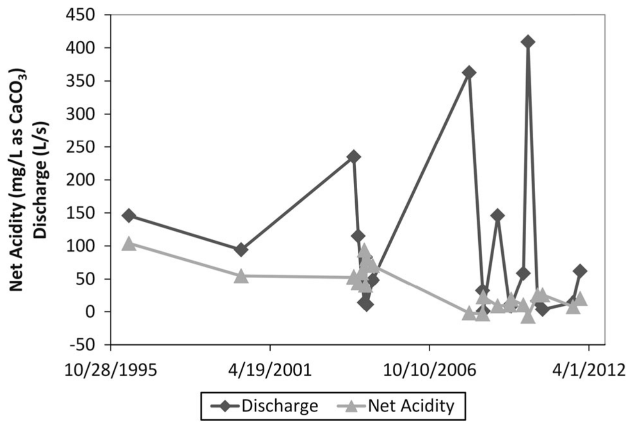

Fig. 6 Discharge (L/s) and net acidity (acidity–alkalinity) (mg/L as $CaCO_3$) over the period of record (1996–2011) for the monitoring site downstream of East Branch Phase 1 reclamation and remediation systems

(Bowman 2011). The project clearly reduced acidity, but given the relationship with drainage, it is difficult to quantify this reduction directly from the raw data (Fig. 6). Analysis by the mean acidity loading method (Fig. 7) shows that, unlike Jobs Hollow, the behavior of East Branch was not affected by reclamation and remediation; both pre-treatment and post-treatment data show a partial flushing model, consistent with the combined effort of land reclamation and water treatment. The linear regressions shown in Fig. 7c are significant (pre-treatment: $p = 1.1 \times 10^{-5}$, post-treatment: $p = 6.5 \times 10^{-5}$). The load reduction is $10^{2.73}$–$10^{1.96}$ or 446 kg/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 (Figs. 8, 9) is illustrative. Data collected at the site of the USGS gauge station at Ewington on Little Raccoon Creek spans nearly four decades (Bowman 2011). The site has a drainage area of 257 km^2 (99.4 square miles). Treatment and reclamation began in Little Raccoon Creek in 1999 and continued through 2006 when the most recent treatment and reclamation project was completed. The data has been split into three time periods representing pre-treatment (1984–1999), treatment in progress (2000–2006), and post-treatment (2007–2013). The analysis shows a shift from a model in which acidity increases at a faster rate than discharge ($F > 1$) for the pre-treatment data, due to the abundance of exposed surface mines and mine spoil, to almost a pure flushing model ($F \approx 1$) once treatment commenced (as shown in the post-treatment data). The treatment and post-reclamation linear regressions are significant ($p = 3.6 \times 10^{-6}$, $p = 7.1 \times 10^{-9}$, respectively), while the pre-treatment linear regression is weaker ($p = 0.17$). The lower confidence level may be due to a limited pre-treatment data set ($n = 4$). The reclamation and treatment projects installed by 1999 lead to an acid load reduction of $10^{4.49}$–$10^{3.08}$ or 29,701 kg/day and the treatment projects installed by

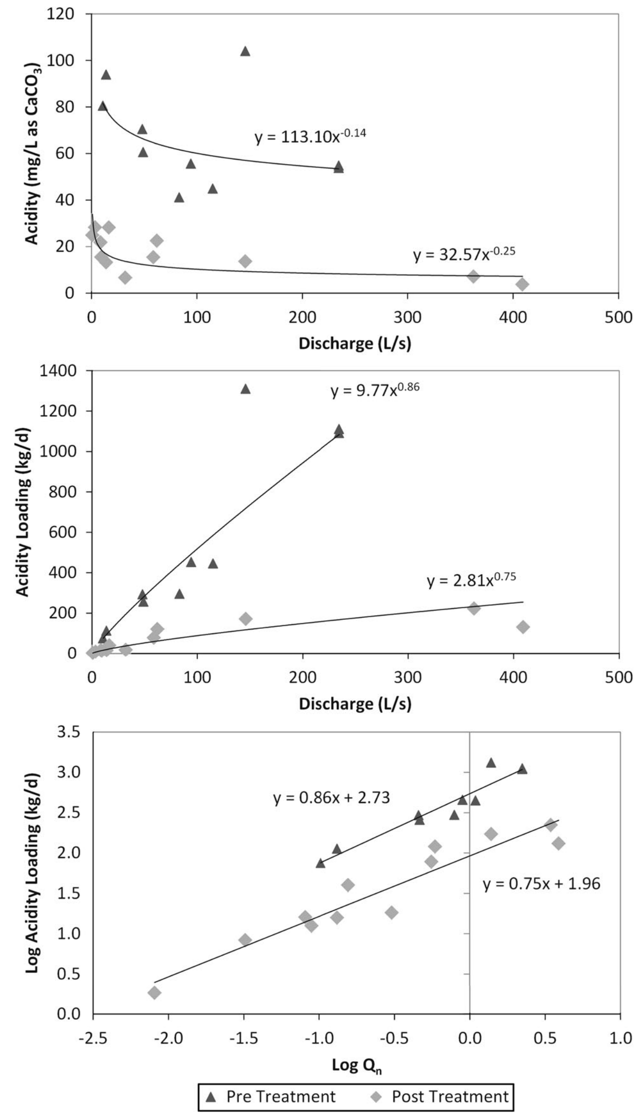

Fig. 7 **a** Total acidity (mg/L as $CaCO_3$) versus discharge (L/s) for pre and post-treatment data sets with power law regressions. **b** Acidity loading (kg/day) versus discharge (L/s) for pre and post-treatment data sets with power law regressions. **c** Log transformed acidity versus log transformed normalized discharge (discharge divided by mean annual discharge) with linear regressions

2006 added an additional $10^{3.08}$–$10^{3.00}$ or 202 kg/day. In total, treatment projects in Little Raccoon Creek have led to a total acid load reduction of 29,903 kg/day.

Method Limitations and Further Applications

The Stoertz's Water Quality Evaluation Method has several practical limitations that should be considered by watershed professionals during data collection and analysis. The first of these is the reliance on a distribution of discharge measurements around the mean annual discharge. It is common to have many measurements at less

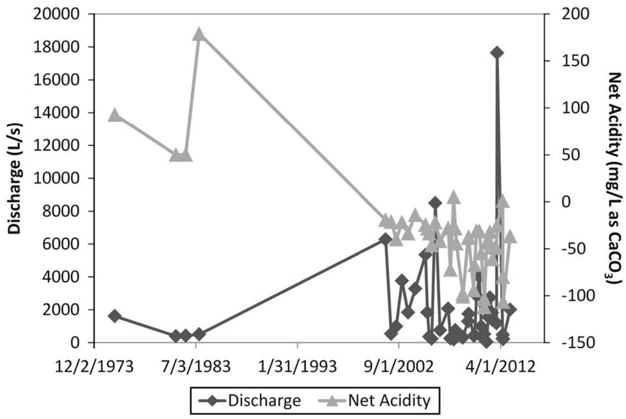

Fig. 8 Discharge (L/s) and net acidity (acidity–alkalinity) (mg/L as $CaCO_3$) over the period of record (1975–2013) for the monitoring site at the USGS gage site on Little Raccoon Creek near Ewington, Ohio

than the mean annual discharge and far fewer at discharge rates higher than the mean annual discharge or very few measurements in total. This measurement bias towards low flow is typically due to the inherent difficulty of measurement and the safety challenge of high flow measurements. The result of the skewed measurement may lead to error in the mean annual acidity loading calculation, although whether it is an overestimation or an underestimation depends on the particular mechanism (e.g. flushing, sparing, dilution) in the case study. When there are limited or no measurements of discharge, methods summarized earlier in this paper may be used to estimate discharge.

Additionally, many mine water treatment systems do not just reduce acidity loading, maintaining some acidity, but add sufficient alkalinity to create net alkaline conditions. Net alkalinity is equivalent to negative acidity, but the nature of the logarithmic scale analysis does not allow for analysis of negative acidity loads. Two methods may be used to combat this: analyzing net alkaline post-treatment data separately from acidic pre-treatment data or adding a 'scalar' to move all values into positive mathematical space. Separating alkaline conditions from acidic conditions into separate analyses will yield an acidity loading at mean annual discharge for the pre-treatment data and a net alkalinity loading at a mean annual discharge for post treatment data. Since alkalinity loading is equal to negative net acidity loading at the same discharge, the results may be summed to determine the total acidity loading reduction. This method is amenable for pre-treatment data that is consistently net acidic and post-treatment data that is consistently net alkaline. Alternatively, a 'scalar' value that is larger than the greatest negative value may be added to all values before log transformation to move all of the data into positive mathematical space. The calculated difference (load reduction) between pre-treatment and post-treatment acidity loading at mean annual discharge will remain the

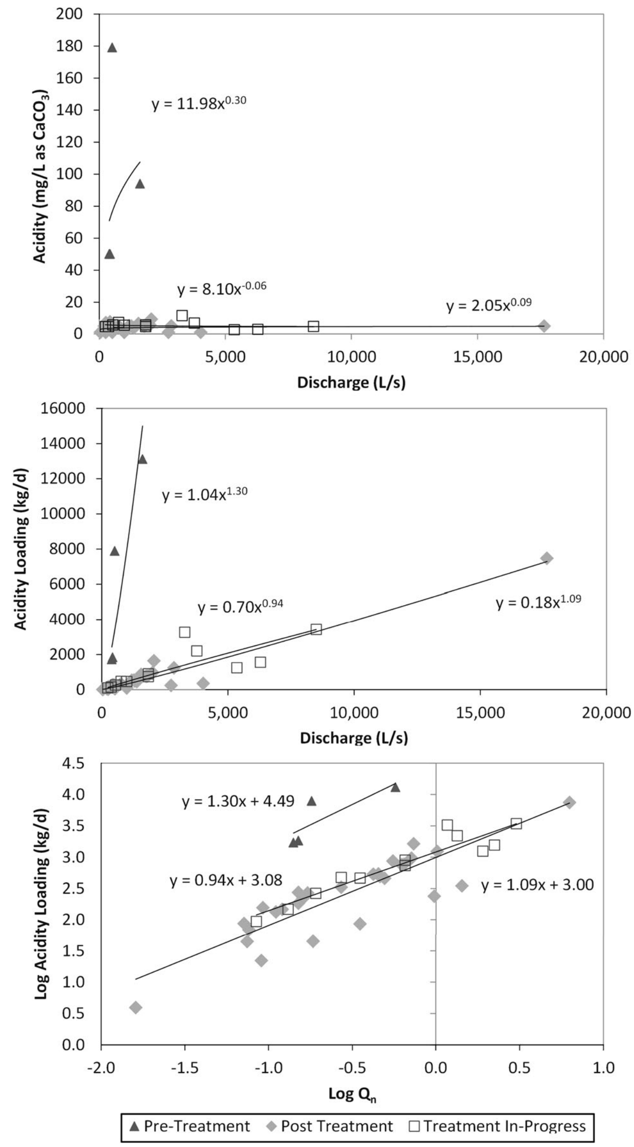

Fig. 9 **a** Total acidity (mg/L as $CaCO_3$) versus discharge (L/s) for pre, in-progress and post-treatment data sets with power law regressions. **b** Acidity loading (kg/d) versus discharge (L/s) for pre, in-progress and post-treatment data sets with power law regressions. **c** Log transformed acidity versus log transformed normalized discharge (discharge divided by mean annual discharge) with linear regressions

same since all values are scaled by the same amount. For reporting acidity loading (rather than load reduction), the scalar value would need to be subtracted from the calculated value.

Where the mean annual discharge is not constant with time, further assumptions must be made. Climate change or major land use change can alter the watershed hydrology and change the mean annual streamflow. Mining can also alter the natural water budget and affect surface and subsurface flow paths. For example, underground mines can facilitate the transfer of surface water to groundwater and the transport of water across watershed boundaries. Stream restoration and mine reclamation can alter the water budget but may not replicate pre-mining conditions (e.g. Cravotta et al. 2013; Goode et al. 2011). Where mined conditions are known to affect the hydrology, drainage areas (and therefore mean annual discharge) can be adjusted accordingly.

Natural attenuation of acidic waters can occur (e.g. Wilkin 2008). However, those processes usually occur at a relatively slow rate, as is clear from the data collected by different watershed groups in SE Ohio over more than 15 years (see http://www.watershedata.com). For that reason, the contribution of natural attenuation to the relatively fast neutralization of remediated sites is probably negligible. If natural attenuation occurs at a faster rate (e.g. exponential decrease), the data could be divided according to the decrease in concentration defined by the decay constant. For example, if the data covers more than a 50 % decrease in concentration, the data could be divided into several periods of time that show a statistically significant decreases in acidity load and the decreases could be compared using the Stoertz Method. However, the decision of how to treat that data will depend on the number of data points and the distribution of values.

The Stoertz's Method has been applied to acidity here. However, the method could be applied to other relatively conservative chemical constituents of interest, such as sulfate, as long as we have sufficient data pre-and post-remediation to be able to obtain the statistically significant regression equations for the logarithmic plots and the flushing factor. Examples of this can be seen in Bowman (2011), which was prepared by one of the authors.

Conclusions

The mean annual load calculated using the Stoertz Method represents 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 Stoertz Method provides two useful pieces of information: the relationship between discharge and loading and the loading at mean annual discharge. The linear relationship between log transformed discharge and loading can be constructed for subsets of data from before and after treatment. Comparing the mean annual loads for each subset yields a performance measure: decrease in mean annual load. The Stoertz Method may be applied by watershed managers as a simple way to measure water quality improvement, as shown in the case studies presented here. It can also be used to evaluate stream degradation, using background data before an activity that can degrade a watershed, such as mining.

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. 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, and habitat quality, will have to be examined.

References

Allard M, Moreau G (1987) Effects of experimental acidification on a lotic macroinvertebrate community. Hydrobiologia 144:37–49

Bonta JV, Dick WA (2003) Impact of coal surface mining and reclamation on surface water chemical concentrations and load rates in three Ohio watersheds. J Am Water Resour As 39(4): 793–815

Bowman J (2011) Nonpoint source monitoring project for acid mine drainage: an evaluation of water quality, biology and acid mine drainage reclamation in five watersheds: Raccoon Creek, Monday Creek, Sunday Creek, Huff Run and Leading Creek. Voinovich School of Leadership and Public Affairs, Ohio Univ, Athens, OH, USA

Clesceri LS, Greenberg AE, Eaton AD (eds) (1998) Acidity/titration method. In: Standard methods for the examination of water and wastewater, 20th edn, part 2310, American Public Health Assoc, American Water Works Assoc, Water Environment Federation, Washington, pp 2-23

Cravotta CA III, Goode DJ, Bartles MD, Risser DW, Galeone DG (2013) Surface-water and groundwater interactions in an extensively mined watershed, upper Schuylkill River. Hydrological Processes, Pennsylvania. http://dx.doi.org/10.1002/hyp.9885

Davis WS, Snyder BD, Stribling JB, Stoughton C (1996) Summary of state biological assessment programs for streams and wadeable rivers. EPA 230-R-96-007, US Environmental Protection Agency Office of Policy, Planning and Evaluation, Washington

DeNicola DM, Stapleton MG (2002) Impact of acid mine drainage on benthic communities in streams: the relative roles of substratum vs. aqueous effects. Environ Pollut 119:303–315

Earle J, Callaghan T (1998) Effects of mine drainage on aquatic life, water uses, and man-made structures. In: Brady KBC, Smith MW, Schueck JH (eds) Coal mine drainage prediction and pollution prevention in Pennsylvania. Pennsylvania Dept of Environmental Protection 5600-BK-DEP2256, Harrisburg, pp 41–410

Gibbons RD (1994) Statistical methods for groundwater monitoring. Wiley, NYC

Goode DJ, Cravotta CA III, Hornberger RJ, Hewitt MA, Hughes RE, Koury DJ, Eicholtz LW (2011) Water budgets and groundwater volumes for abandoned underground mines in the Western Middle Anthracite Coalfield, Schuylkill, Columbia, and Northumberland Counties, Pennsylvania—preliminary estimates with identification of data needs. USGS Scientific Investigations Report 2010–5261, Harrisburg

Hedin RS (2004) The use of measured and calculated acidity values to improve the quality of AMD data sets. In: Proceedings of American Society for Mining and Reclamation Conference, Morgantown, p 879–898

Hedin RS, Nairn RW, Kleinmann RLP (1994) Passive treatment of coal mine drainage. Bureau of Mines Information Circular 9389, United States Department of the Interior, Washington, DC

Helsel DR, Hirsch RM (2002) Statistical methods in water resources. USGS Techniques of Water Resources Investigations 04-A3, Washington

Hirsch RM, Slack JR, Smith RA (1982) Techniques of trend analysis for monthly water quality data. Water Resour Res 18(1):107–121

Johnson NM, Likens GE, Bormann FH, Fisher DW, Pierce RS (1969) A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire. Water Resour Res 5(6):1353–1363

Karr JR (1981) Assessment of biotic integrity using fish communities. Fisheries 6(6):21–27

Karr JR, Yoder CO (2004) Biological assessment and criteria improve total maximum daily load decision making. J Environ Eng 130(6):594–604

Kendall MG (1975) Rank correlation methods, 4th edn. Charles Griffin, London

Kirby CS, Cravotta CA III (2005a) Net alkalinity and net acidity 1: theoretical considerations. Appl Geochem 20:1920–1940

Kirby CS, Cravotta CA III (2005b) Net alkalinity and net acidity 2: practical considerations. Appl Geochem 20:1941–1964

Koltun GF, Whitehead MT (2002) Techniques for estimating selected streamflow characteristics of rural, unregulated streams in Ohio. USGS Report FHWA/OH-2001/13, Columbus

Kowalik RA, Cooper DM, Evans CD, Ormerod SJ (2007) Acidic episodes retard the biological recovery of upland British streams from chronic acidification. Glob Chang Biol 13(11):1365–2486

Kruse NA, Bowman JR, Mackey AL, McCament B, Johnson KS (2012) The lasting impacts of offline periods in lime dosed streams: a case study in Raccoon Creek, Ohio. Mine Water Environ 31(4):266–272

Lacroix GL (1987) Fish community structure in relation to acidity in three Nova Scotia rivers. Can J Zool 65(12):2908

Lewis WM, Grant MC (1979) Relationships between stream discharge and yield of dissolved substances from a Colorado mountain watershed. Soil Sci 128(6):353–363

Liebermann TD, Middelburg RF, Irvine SA (1987) User's manual for estimation of dissolved-solids concentrations and loads in surface water. USGS WRI Report 86-4124, Denver

López DL, Stoertz MW (2001) Chemical and physical controls on waters discharged from abandoned underground coal mines. Geochem- Explor Env A 1:51–60

Mann HB (1945) Nonparametric tests against trend. Econometrics 13:245–259

Montgomery DC, Johnson LA (1976) Forecasting and time series analysis. McGraw-Hill, NYC

Watzlaf GR, Schroeder KT, Kleinmann RLP, Caries CL, Nairn, RW (2004) The passive treatment of coal mine drainage. US Dept of Energy NETL-2004/1202. http://www.netl.doe.gov/technologies/coalpower/ewr/water/pdfs/Passive%20Treatment.pdf. Accessed Nov 2013

NRC (National Research Council) (2001) Assessing the TMDL approach to water quality management. National Academy Press, Washington

OEPA (1988) Biological criteria for the protection of aquatic life: vol I: the role of biological data in water quality assessment. Ohio Environmental Protection Agency, Columbus

Ohio Environmental Protection Agency (OEPA) (1987) Biological criteria for the protection of aquatic life: vol II: users manual for biological field assessment of Ohio surface waters. Ohio Environmental Protection Agency, Columbus

Pigati E, López DL (1999) Effect of subsidence on recharge at abandoned coal mines generating acidic drainage: the Majestic Mine, Athens County, Ohio. Mine Water Environ 18(1):45–66

Ries KG III, Guthrie JD, Rea AH, Steeves PA, Stewart DW (2008) StreamStats—a water-resources web application. USGS Fact Sheet 2008–3067, Washington

Rosemond AD, Reice SR, Elwood JW, Mulholland PJ (1992) The effects of stream acidity on benthic invertebrate communities in the south-eastern United States. Freshwater Biol 27:193–209

Searcy JK, Hardison CH (1960) Double-mass curves. USGA Water Supply Paper 1541-B, Washington

Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. J Am Stat As 63:1379–1389

Skousen JG, Ziemkiewicz PF (1996) Acid mine drainage control and treatment, 2nd edn. West Virginia Univ and the National Mine Land Reclamation Center, Morgantown

Smith RA, Hirsch RM, Slack JR (1982) A study of trends in total phosphorus measurements at NASQAN stations. USGS Water Supply Paper 2190, Alexandria

Stoertz MW, Green DH (2007) A practical measure of acidity loading in coal-mined watersheds. Abstracts with Programs, Geological Soc of America Annual Meeting, Denver 39(6):134

Stoertz MW, Hughes ML, Wanner NS, Farley ME (2001) Long-term water quality trends at a sealed, partially flooded underground mine. Environ Eng Geosci VI I(1):51–65

Stoertz M, Bourne H, Knotts C, White M (2002) Effects of isolation and acid mine drainage on fish and macroinvertebrate communities of Monday Creek, Ohio. Mine Water Environ 21(2):60–72

USEPA (2013a) Summary of the Clean Water Act. http://www2.epa.gov/laws-regulations/summary-clean-water-act. Accessed June 7, 2013

USEPA (2013b) National Rivers and Streams Assessment 2008–2009 A Collaborative Survey. US Environmental Protection Agency Office of Wetlands, Oceans and Watersheds Draft Report EPA/841/D-13/001, Washington DC, USA

USEPA (US Environmental Protection Agency) (2012) Overview of Impaired Waters and Total Maximum Daily Loads Program. http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/intro.cfm. Accessed June 7, 2013

Wilkin RT (2008) Contaminant attenuation processes at mine sites. Mine Water Environ 27:251–258

Ziemkiewicz PF, Skousen JG, Simmons J (2003) Long-term performance of passive acid mine treatment systems. Mine Water Environ 22(3):118–130