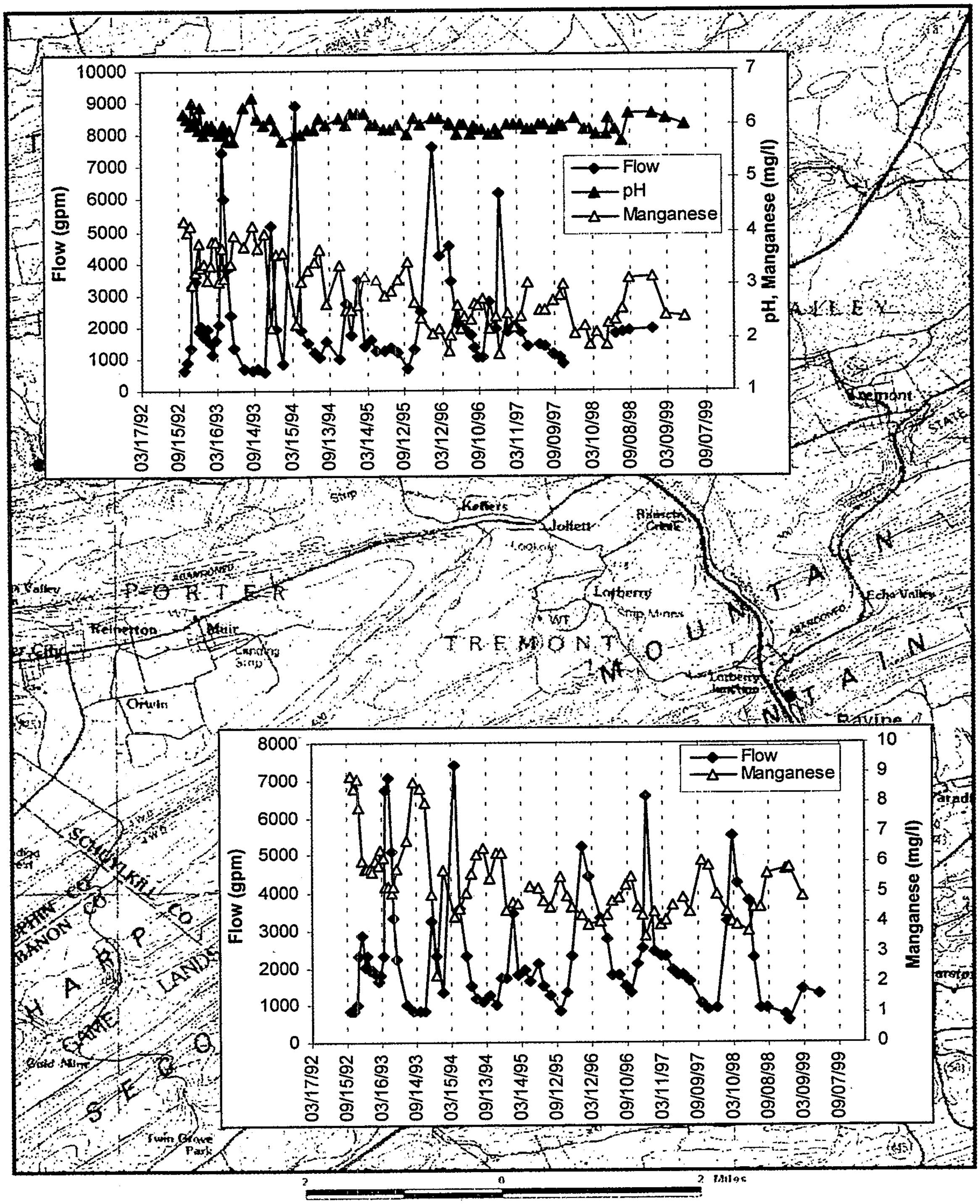
United States
Environmental Protection
Agency

Office of Water
(4303)

EPA 821-B-01-011
December 2001
FINAL

EPA

Coal Remining Statistical Support Document



EPA-821-B-01-011

COAL REMINING
STATISTICAL SUPPORT DOCUMENT

DECEMBER 2001

Office of Water
Office of Science and Technology
Engineering and Analysis Division
U.S. Environmental Protection Agency
Washington, DC 20460

Acknowledgments

This document was developed under the direction of William A. Telliard and John F. Fox of the Engineering and Analysis Division (EAD) within the U.S. Environmental Protection Agency's (EPA's) Office of Science and Technology (OST). EPA gratefully acknowledges Roger Hornberger, Michael W. Smith, Daniel J. Koury, and J. Corey Cram of Pennsylvania's Department of Environmental Protection (PADEP) for their contributions to this document. EPA also wishes to thank DynCorp Information and Enterprise Technology for its invaluable support in developing this document.

Disclaimer

The primary contact regarding questions or comments on this document is:

John F. Fox
Engineering and Analysis Division (4303)
U.S. Environmental Protection Agency
Ariel Rios Building, 1200 Pennsylvania Avenue, N.W.
Washington, DC 20460
Phone: 202/260-9889
email: fox.john@epa.gov

TABLE OF CONTENTS

List of Tables

List of Figures

Section 1.0 Introduction

Section 2.0 Characteristics of Coal Mine Drainage Discharges

Section 3.0 Statistical Methodology for Establishing Baseline Conditions and Setting Discharge Limits at Remining Sites

Section 4.0 Baseline Sampling Duration and Frequency

Section 5.0 Long-term Monitoring Case Studies

Section 5.0 Long-term Monitoring Case Studies (cont.)

Section 1.0 Introduction

Acid mine drainage has been produced by coal mining operations in the Appalachian Coal Region of the eastern United States and elsewhere for many years, resulting in extensive surface-water and ground-water pollution. The Federal Clean Water Act (CWA), the Federal Surface Mining Control and Reclamation Act (SMCRA), and associated state laws require coal mine operators to take steps to prevent or control the production of acid mine drainage, and to treat acid mine drainage from active and reclaimed surface mining operations so that point-source discharges meet the applicable effluent limitations found at 40 CFR part 434.

Much of the acid mine drainage occurring in the Appalachian Coal Region is emanating from abandoned surface and underground mines that were mined and abandoned prior to the enactment of SMCRA and the CWA. According to the Appalachian Regional Commission (1969), 78 percent of the acid mine drainage produced in northern Appalachia is associated with inactive or abandoned mines. More recent U.S. Geological Survey reports (Wetzel and Hoffman, 1983, 1989) provide summaries of surface-water quality data and patterns of acid mine drainage problems throughout the Appalachian Coal Basin. A set of companion reports (Hoffman and Wetzel, 1983, 1989) contain similar information for the Interior Coal Province of the Eastern Coal Region of the United States. Current EPA data document that the number one water quality problem in Appalachia is drainage from abandoned coal mines, resulting in over 9,700 miles of acid mine drainage polluted streams. A 1995 EPA Region III survey found that 5,100 miles of streams in four Appalachian states are impacted by acid mine drainage, predominantly from abandoned coal mines. Pennsylvania alone accounts for approximately 2,600 acid mine drainage impacted stream miles.

The remaining coal reserves in these abandoned mine land areas frequently make them attractive for the active mining industry; but traditionally, potential liability for the treatment of the abandoned mine drainage established a disincentive to the permitting and remining of these areas. If a pre-existing pollutional discharge of acid mine drainage was occurring within the area or on an area hydrologically connected to the permit area, mine operators often faced liability to treat the discharge to best available technology economically achievable (BAT) effluent standards (40 CFR part 434).

1.1 Remining Program History

In the 1980s, changes to the CWA (see 1987 Amendment to Section 301; the Rahall Amendment) and state mining laws (e.g., 1984 Amendment to the Pennsylvania Surface Mining Conservation and Reclamation Act (PA SMCRA)) provided incentives to mine operators to remine areas with pre-existing pollutional discharges. Pursuant to these changes in state and federal laws, the flow and water quality characteristics or "baseline pollution load" of these pre-existing discharges must be documented prior to the commencement of the remining operation. Under this program, the mine operator submits a surface mining permit application including: (1) sufficient baseline pollution load data, and (2) a pollution abatement plan which demonstrates how the remining operation proposes to eliminate or reduce the pre-existing pollution. The regulatory authority completes a "best professional judgement (BPJ) analysis" pursuant to Section 402(a) of the CWA as part of the permit review process. A BPJ-based remining permit may be issued that requires the mine operator to treat the pre-existing discharges only if the baseline pollution load has been exceeded, and then only treat the discharges to baseline pollution load levels rather than to conventional BAT effluent standards. The procedures to determine the level of treatment required to meet baseline is not standard, and is dependent on various site-specific elements of the BPJ.

BPJ is defined as: "The highest quality technical opinion forming the basis for the terms and conditions of the treatment level required after consideration of all reasonably available and pertinent data. The treatment levels shall be established in accordance with Sections 301 and 402 of the Federal Water Pollution Control Act (33 USC §§1311 and 1342)." BPJ-determined effluent limits must be based upon BAT or any more stringent limitation necessary to ensure the discharge does not violate state in-stream water quality standards. Theoretically, BPJ-determined treatment levels can range from the pre-existing baseline level to the conventional BAT limits.

The BPJ analysis is a BAT analysis in miniature, specific to an individual mine site, rather than an entire class of industrial wastewater discharges (i.e., surface coal mining). For a remining permit, the analysis should consider the cost of treatment to conventional surface mining BAT levels, as well as the cost of achieving pollution load reduction through the implementation of a pollution abatement plan. The permit writer also should consider any unique factors pertaining to the proposed remining operations and any potential adverse or beneficial non-water quality environmental impacts.

The Rahall Amendment to the Federal Clean Water Act in 1987 provided a foundation for the development of effective remining programs in many coal mining states. Between 1984 and 1988, EPA and the Commonwealth of Pennsylvania Department of Environmental Protection (PADEP) cooperated in a remining project. The purpose of that project was to develop an effective remining program pursuant to the 1984 Amendments to PA SMCRA, that would not be in conflict with the provisions of the Federal Clean Water Act and the associated 40 CFR part 434 regulations. Pennsylvania promulgated remining regulations on June 29, 1985, that were approved by the U.S. Office of Surface Mining Reclamation and Enforcement (OSMRE) on February 19, 1986.

The work products of the PADEP/EPA cooperative study included preliminary treatment costing and remining costing studies in 1986 (prepared by Kohlmann Ruggiero Engineers (KRE), and Phelps and Thomas of the Pennsylvania State University), the development of the REMINE computer software package and Users Manual in 1987, and associated technical reports in 1987 and 1988. Included in these technical reports was the final treatment costing study (Kohlmann Ruggiero Engineers, P.C., 1988) and a series of eight water quality statistical reports prepared by Dr. J. C. Griffiths of the Pennsylvania State University. Since these unpublished statistical analyses of mine drainage datasets are relevant to baseline pollution load statistics, they are presented in an abridged form in a companion volume to this report, prepared by US EPA (2001; EPA-821-B-01-014).

1.2 Pennsylvania DEP Remining Permitting Procedures

Since 1985, PADEP has issued approximately 300 remining permits, with a 98 percent success rate. A successful remining site is one that has been mined without incurring treatment liability as the result of exceeding the baseline pollution load of the pre-existing discharges. Data from 112 of these sites that have been completely reclaimed have been used by EPA in evaluating Best Management Practice (BMP) performance (see EPA Coal Remining BMP Guidance Manual). The elements of establishing baseline pollution loads and measuring compliance that are provided in the Pennsylvania program guidance on the BPJ process are summarized below.

1.2.1 Pre-existing Discharges

Various relationships exist between the permit boundaries of surface coal mine sites and the location of pre-existing pollutional discharges. The simplest relationship exists where a single abandoned mine drainage discharge point is located within, or closely adjacent to, the proposed surface mine permit boundaries, and the proposed mine is the only active mining operation. A more complex relationship occurs where numerous pre-existing discharges from the same coal seam, or from multiple coal seams, are located within the proposed surface mine permit area or are hydrologically connected to that permit area. In addition, there are situations where more than one active mining operation is hydrologically connected to the same pre-existing discharge. The PADEP program includes considerations for: (a) monitoring and baseline data collection of single and multiple discharges, (b) establishing baseline pollution load of single or multiple discharges through statistical methods, and (c) determining compliance with BPJ determined effluent limits for the pre-existing discharges through statistical methods and permit conditions (for individual operators and multiple operators).

In many cases, pre-existing pollutional discharges may occur in the form of numerous discharge points, all of which emanate from a hydrologically discrete ground-water flow system. Ground-water flow paths may change during and following remining such that new discharge points appear, former discharge points disappear, and/or the distribution of flow rates between discharges changes. Where this situation is likely to occur, it is usually advantageous to designate hydrologic units. Each unit must be a hydrologically discrete area such that ground water from one hydrologic unit does not flow to a different hydrologic unit. Hydrologic unit boundaries must be determined for situations where two or more discharges are to be aggregated for load calculations.

Discharges may be combined either naturally or by man-made controls to a single monitoring point, provided that the combination of discharges does not affect the pollution load measurement and that discharges from different hydrologic units are not combined. It is usually desirable (in terms of cost to the operator, permit writing, and compliance monitoring) for the permit applicant to minimize the number of monitoring points needed.

The permit applicant must perform a baseline pollution load statistical determination for each monitoring point. Where multi-discharge hydrologic units are defined, the baseline statistics should be calculated for the aggregate pollution load from the hydrologic unit. That is, loads are summed for all the discharges in the hydrologic unit on a given date. Baseline pollution load determination of a hydrologic unit requires sampling and analysis of each discharge on the same date using an equal number of samples from each discharge. The baseline pollution load is then reported as the combined pollution load from the hydrologic unit.

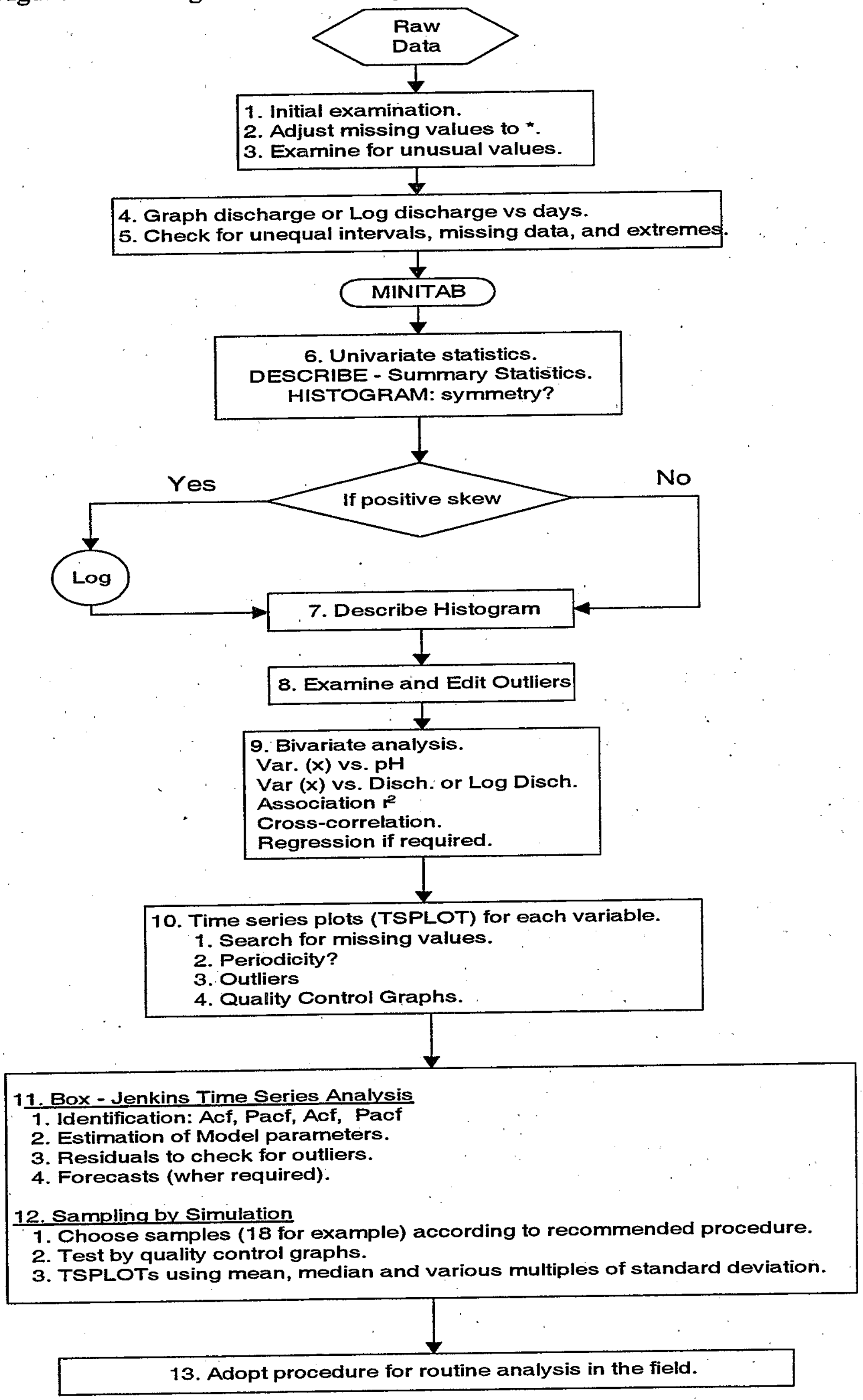
1.2.2 Baseline Pollution Load - Determination and Compliance Monitoring

The process of establishing a realistic baseline pollution load for a mine site requires knowledge of hydrology and statistics. An adequate number of samples must be collected at sufficient time intervals to represent seasonal variations throughout the water year (October through September). The statistical components of establishing baseline pollution load include characterizing the patterns of variation and measuring central tendency, so that any mining-induced changes in pollution load can be distinguished from seasonal and random variations. During active mining and post-mining, individual sampling events and the statistical summary of data collected over successive water years are compared to the pre-mining baseline statistics.

An algorithm for the statistical analysis of mine drainage discharge data was developed by Griffiths (1987) for use in the Pennsylvania program and elsewhere (Figure 1.2a). The algorithm included a simple quality control approach using the exploratory data analysis methods developed by Tukey (1977), and used bivariate statistical methods and time series analyses, where appropriate (e.g., research purposes documented in eight statistical reports by Griffiths, 1987 and 1988). In practice, almost all of the remining permits issued under the Pennsylvania program have used the Baseline Pollution Load Statistical Results Summary presented in Table 1.2a. The five statistical calculations (range, median, quartiles, 95 percent confidence interval about the median, and 95 percent tolerance interval) are based upon Tukey's exploratory data analysis methods and order statistics. Alternative statistical calculations may be used in place of the calculations identified on Table 1.2a, provided that the permit applicant demonstrates that the alternative calculations are statistically valid and applicable. For example, the mean and variance may be used if the data are normally distributed. The REMINE computer software package developed by EPA, PADEP and the Pennsylvania State University was integrated with the MINITAB statistical software package[1], which includes statistical and graphical methods to perform all of the steps in the algorithm presented in Figure 1.2a.

[1]MINITAB is a commercial software package from Minitab, Inc., © 1986, 3081 Enterprise Drive, State College, PA 16801

Figure 1.2a: Algorithm for Analysis of Mine Drainage Discharge Data



While the permit applicant is responsible for submitting the baseline pollution load data and statistical summary, the permit reviewer must check the calculations to ensure that the results are correct. In addition, the reviewer must examine the distribution of the data to determine whether a logarithmic transformation is appropriate. If logarithmic transformation results in a more normal distribution curve, log-transformed data should be used in determining the baseline.

Each discharge point or hydrologic unit will have a baseline pollution load summary. Compliance of discharge points is determined by comparing monthly sample analysis results to the determined baseline pollution load. A discharge point is considered to be in compliance as long as the sample analysis indicates that the pollution load does not exceed either the 95 percent tolerance limit (item 4, Table 1.2a) or the 95 percent confidence interval about the median (item 5, Table 1.2a). The confidence intervals around the median are calculated using the equation noted in Table 1.2a, and taken from McGill, Tukey, and Larsen (1978).

Table 1.2a: Baseline Pollution Load Summary

Mine ID: ______ Mine Name: ____________________________ Hydrologic Unit ID:_____

# of Samples:____		Flow	Loading in Pounds Per Day				
Statistical Results		(gpm)	Acidity	Iron	Manganese	Aluminum	Sulfates
1. Range	Low:						
	High:						
2. Median							
3. Quartiles	Low:						
	High:						
4. Approximate 95% tolerance limits	Low:						
	High:						
5. 95% Confidence Int. about median*	Low:						
	High:						

***Note:** Confidence intervals about median = M +/- 1.58[1.25R/(1.35xSQR(N)] where:

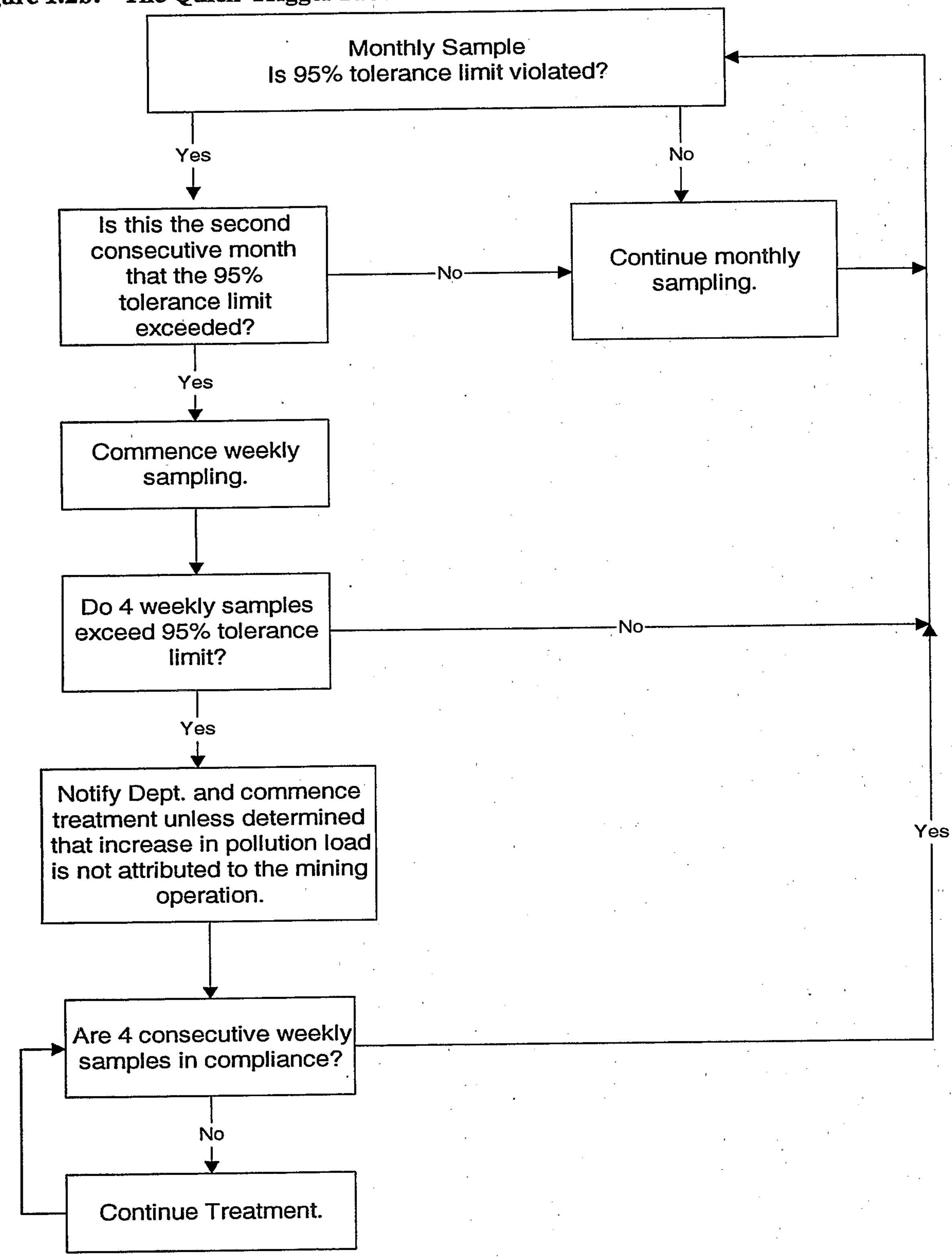
M = median, R = range between quartiles, and SQR(N) = the square root of the number of samples (McGill et. al., 1978).

An excursion (i.e., apparent violation) of the baseline pollution load occurs when the result of a sample analysis exceeds the 95 percent tolerance limit, or when a median of sample results obtained over a new water year is outside the bounds of the 95 percent confidence interval about the original baseline median. An excursion of the baseline pollution load above the 95 percent tolerance limit is known as a "quick trigger" violation. A more subtle and long-term trend of the pollution load above the 95 percent confidence interval about the median is known as a "subtle trigger" violation.

The 95 percent tolerance limit is determined by ranking the data in order of magnitude and dividing the data into 32 increments. The 95 percent tolerance limits correspond to the lowest and highest of the 32 increments. For datasets containing 16 or fewer samples, the approximate 95 percent tolerance limits correspond to the smallest and largest sample values. The upper 95 percent tolerance limit is the "quick trigger" or "critical" value mechanism for monthly monitoring data. If two consecutive monthly samples exceed the upper 95 percent tolerance limit, weekly monitoring is initiated. The quick trigger values are provided in the Surface Mine Permit. Permit conditions specify quick trigger monitoring and compliance steps shown in Figure 1.2b.

Determination of long-term compliance with the subtle trigger typically involves comparison of the pollution loading data for successive water years to the 95 percent confidence limit about the median for the baseline. The 95 percent confidence limits are also based on baseline data (Table 1.2a) and given in the Surface Mine Permit. The 95 percent confidence interval is defined as the range of values around the median in which the true population median occurs with a 95 percent probability. This value is used to determine if statistically significant changes in median pollution loads have occurred between the baseline monitoring period and water years during mining and postmining. Permit conditions specify the process that will be used to determine compliance with the subtle trigger.

Figure 1.2b: The Quick-Trigger Process



Box plots can be used to easily compare the baseline pollution load (or concentration) for different analytes to successive water-year datasets. Box plots can be constructed to show the median value, the 95 percent confidence limits around the median and the upper and lower quartiles and range of data. The length of the box corresponds to the interquartile range (IQR) equal to the 75th percentile minus the 25th percentile. Therefore, 50 percent of the data will fall within the range given by the length of the box. The upper whisker (the t-shaped line above the upper end of the boxes) extends to the largest value less than or equal to the 75^{th} percentile plus 1.5 times the IQR. Likewise, the lower whisker extends to the smallest value greater than or equal to the 25^{th} percentile minus 1.5 times the IQR. Any value that is beyond the whiskers is known as an extreme value. Extreme values less than 1.5 IQRs away from the nearer whisker (or equivalently, less than 3 IQRs away from the edge of the box) are represented by an open circle. Extreme values beyond 1.5 IQRs away from the nearer whisker are represented by an "x".

Figures 1.2c and 1.2d are examples of box plots of water quality data from the Dunkard Creek in Greene County, Pennsylvania (this acid mine drainage impacted stream segment is featured in several other figures in Section 2.0). Figure 1.2c shows variations in the range, median and quartiles of pH distributions for three time periods corresponding to significant changes in mining regulation and acid mine drainage (AMD) control and abatement in Pennsylvania. The box plot of pH data from 1950 to 1965 represents all available data (N=54) at this monitoring point prior to the Pennsylvania Clean Streams Law requiring that active mines treat acid mine drainage. This law went into effect in 1966. The box plot of pH data from 1983 to 1997 (N=175) represents the time period following the approval of Pennsylvania for primacy to regulate the Federal SMCRA of 1977. Primacy provided for significant increases in staff and resources for permitting, inspection and enforcement of active mine sites, and funds to reclaim abandoned mine sites with AMD problems. The median pH of 6.9 for the data (N=112) for the intermediate time period (1966-1982) is significantly different than the median pH of 3.95 for the time period prior to the 1966 AMD treatment requirement. Figure 1.2d shows box plots of manganese concentrations for the same monitoring point and same time periods as Figure 1.2c. The median and interquartile range for the 1966-1982 data (N=107) is significantly less than that of the 1950-1965 data (N = 14); and the range of the manganese concentrations in the 1983-1997 data (N = 173) is less than half of the range for the two previous time periods. Additional examples of box plots and explanations of their origin and development are contained in Tukey (1977), McGill, Tukey, and Larsen (1978), Veleman and Hoaglin (1981) and Helsel (1989).

Figure 1.2c: Dunkard Creek pH

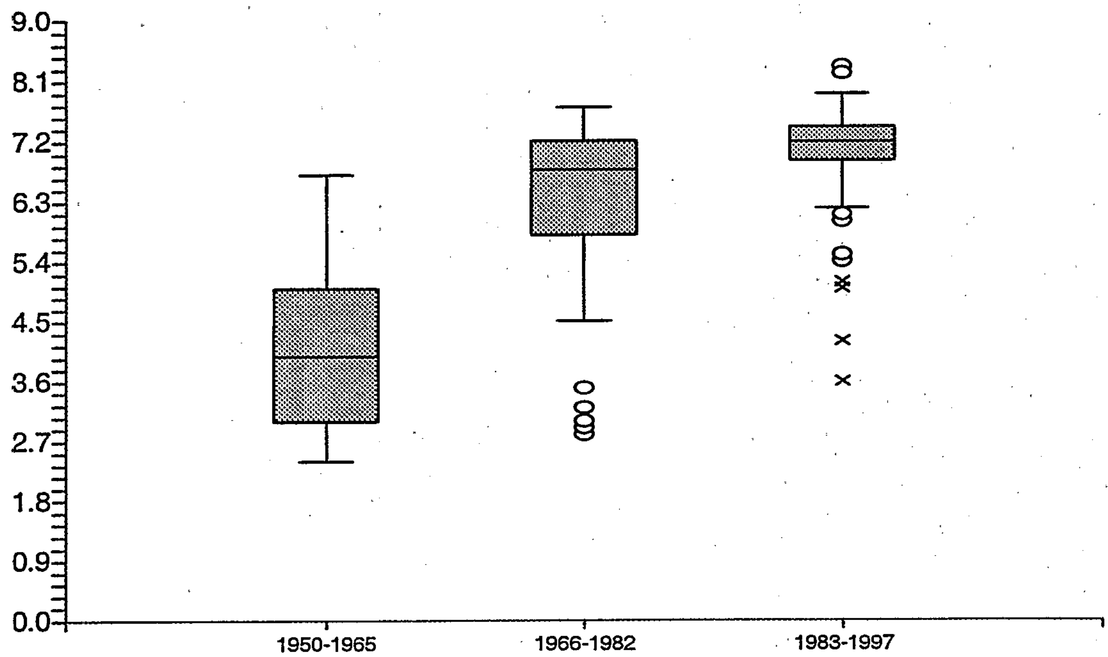
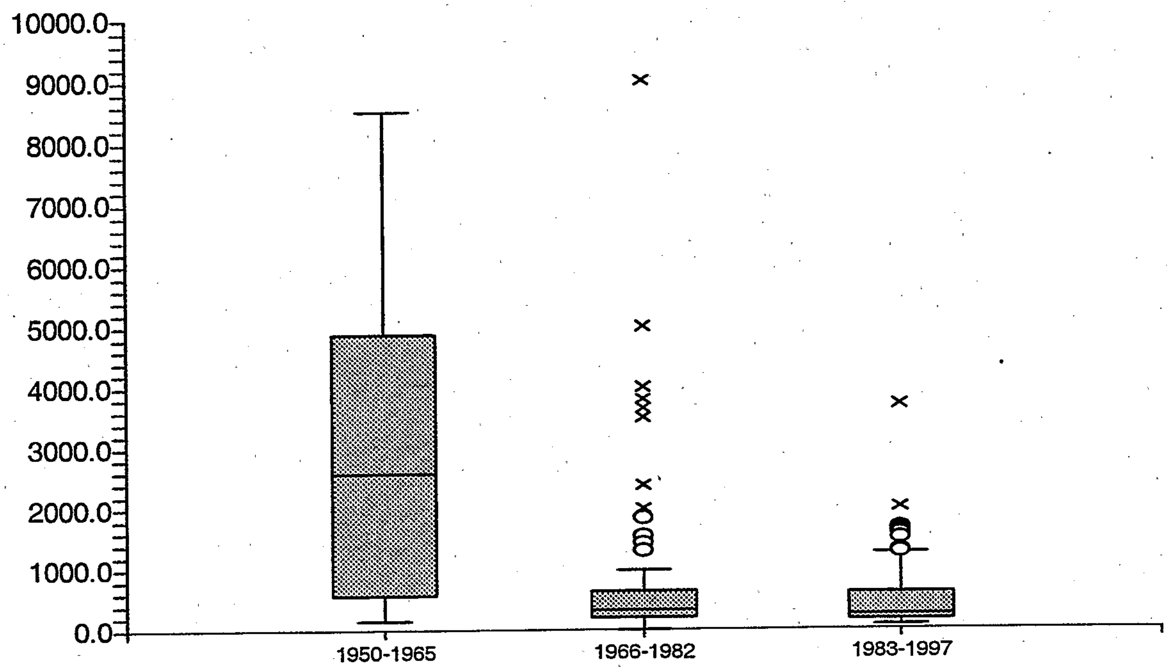


Figure 1.2d: Dunkard Creek Manganese (mg/L)



Monitoring and compliance inspections are conducted periodically (i.e., quarterly). Reviews of the monthly monitoring data for the purpose of comparing current data to the baseline data, checking for subtle and quick trigger violations, and noting any data trends, are conducted annually.

1.3 IMCC Evaluation of State Remining Programs

The Interstate Mining Compact Commission (IMCC) is a multi-state governmental organization representing the natural resource and environmental protection interests of its member states, including extensive interaction with EPA, OSMRE, and other federal agencies. The IMCC organized a national remining task force in 1996, with representation from EPA, OSMRE and member states, in order to develop and promote various remining incentives that would accomplish significant abandoned mine land reclamation and associated water quality benefits. A product of the IMCC Remining Task Force is a discussion paper on water quality issues related to coal remining, for which EPA, OSMRE, and IMCC jointly solicited comments in February 1998 from a wide range of environmental, industry, and government agency commentors/respondents.

In a related effort that is part of a cooperative project with EPA and OSMRE, the IMCC solicited and compiled responses from 20 states on their remining program experiences. This compilation of responses provides extensive information on the number of Rahall-type remining permits issued in the states, the contents of these permits (including the availability of baseline pollution load analyses and data), and the types and effectiveness of BMPs employed during remining operations. A summary of these responses is included as Appendix C of EPA's Coal Remining BMP Guidance Manual. IMCC also submitted 61 data packages to EPA from 6 member states. These data packages include pre-, during, and post-mining water quality data, BMP implementation plans, remining operation plans, geology and overburden analysis data, abandoned mine land conditions, and topographic maps.

Based upon review of the IMCC solicitation responses, 61 data packages, and discussions with state agency representatives, it is evident that baseline pollution load data requirements vary widely from state to state. Pennsylvania, Virginia and some other states generally require a minimum of 12 monthly samples of pre-existing pollutional discharges to calculate the baseline pollution load and characterize seasonal variations throughout the water year. One state water quality agency has advocated the use of 52 weekly samples to characterize baseline pollution load, which may have been a disincentive to remining. That state has only been able to issue a handful of Rahall remining permits. At the other end of the scale, another state has developed a draft sampling protocol to establish baseline pollution load with only 6 monthly samples, similar to the background sampling requirement for determining "probable hydrologic consequences" in most state surface mining permits. The draft protocol divides the water year into high-flow, low-flow, and intermediate-flow periods, and contains requirements for sampling each of these periods and considering transition periods and other hydrologic factors.

The establishment of the baseline pollution load is largely a statistical exercise because many pre-existing discharges are known to be highly variable in flow and/or water quality. The use of statistics is necessary to quantify these variations and to summarize the behavior of the discharges, which may be related to seasonal variations or other hydrologic factors. Statistical analysis of the baseline pollution load data also enables a distinction to be made, after remining has commenced, between normal seasonal variations and mining-induced changes in pollution load that will require initiation of treatment.

The baseline must consist of an adequate number of samples of sufficient intervals and duration, in order to provide adequate protection for both the industry and the regulatory authority against false triggers. The greater the number of samples and range of hydrologic conditions represented by the baseline pollution load determination, the greater the likelihood that the baseline pollution load determination is statistically and hydrologically sound. In attempting to establish the baseline pollution load with a relatively small number of samples, there is an inherent risk of under representation. In establishing national standards or guidelines for baseline pollution load, careful consideration must be given to determining the optimum number of samples, the associated time intervals, and sampling duration in order to achieve statistical and hydrologic credibility, without being overly burdensome, costly, or impractical.

References

Appalachian Regional Commission, 1969. Acid Mine Drainage in Appalachia. Appalachian Regional Commission: Washington, DC.

Griffiths, J.C., (No date). Report No. 1: Reconnaissance into the use of MINITAB using Hamilton 01, 08 Data. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987. Report No. 2: Time Series Analysis of Data from Hamilton 08. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987. Report No. 3: Analysis of Data from Arnot 001, 003, 004. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987. Report No. 4: Analysis of Data from the Clarion Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987. Report No. 5: Analysis of Data from Ernest Refuse Pile, Indiana County, Pennsylvania. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987. Report No. 6: Analysis of Data from the Fisher Deep Mine. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988. Report No. 7: Analysis of Data from the Markson Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988. Report No. 8: Synopsis of Reports 1 - 7. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Helsel, D., 1989. Boxplots: A Graphical Method for Data Analysis. US Geological Survey, Bureau of Systems Analysis Technical Memorandum No. 89.01: Reston, VA.

Hoffman, S.A. and K.L. Wetzel, 1983. Summary of Surface Water Quality Data, Interior Coal Province, Eastern Region, October 1978 to September 1982. US Geological Survey Open File Report 83-941: Harrisburg, PA.

Hoffman, S.A. and K.L. Wetzel, 1989. Distribution of Water Quality Indicators of Acid Mine Drainage in Streams of the Interior Coal Province, Eastern Coal Region of the United States. US Geological Survey Water Resources Investigations Report 89-4043: Harrisburg, PA.

Kohlmann Ruggiero Engineers, P.C., 1988. Development of a Wastewater Treatment Cost Analysis Module for the Coal Remining BPJ Computer Software Package (Preliminary Draft). U.S. Environmental Protection Agency: Washington, DC.

McGill, R., J.W. Tukey, and W.A. Larsen, 1978. Variations of Box Plots. The American Statistician, Vol. 32, No. 1, pp. 12 - 16.

Phelps, L.B. and M.A. Thomas, 1986. Assessment of Technological and Economical Feasibility of Surface Mine Reclamation, Stage One Research. Pennsylvania Department of Environmental Resources Contract No. 898621: Harrisburg, PA.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison Wesley Publishing Company: Reading, MA.

U.S. EPA, 1986. Preliminary Engineering Cost Manual for Development of BPJ Analyses for Coal Remining, Four Case Histories in Pennsylvania. Prepared for U.S. Environmental

Protection Agency, Office of Water, Industrial Technology Division Energy and Mining Branch, Washington D.C. by Kohlmann Ruggiero Engineers, P.C.

U.S. EPA, 2001. Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits. Prepared for US Environmental Protection Agency, Office of Water by Pennsylvania Department of Environmental Protection and DynCorp I&ET. EPA-821-B-01-014.

Velleman, P.F. and D.C. Hoaglin, 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press: Boston, MA.

Wetzel, K.L. and S.A. Hoffman, 1983. Summary of Surface Water Quality Data, Eastern Coal Province, October 1978 to September 1982. US Geological Survey Open File Report 83-940: Harrisburg, PA.

Wetzel, K.L. and S.A. Hoffman, 1989. Distribution of Water Quality Characteristics that May Indicate the Presence of Acid Mine Drainage in the Eastern Coal Province of the United States. US Geological Survey Hydrologic Investigations Atlas HA-705: Harrisburg, PA.

Section 2.0 Characteristics of Coal Mine Drainage Discharges

Acid mine drainage is generated when sulfide minerals, principally pyrite (FeS_2), are exposed to increased amounts of air and water in the oxidizing and non-alkaline environment of a surface or underground mine. The sulfide minerals typically occur in coal beds as well as in strata overlying and underlying the coal. Weathering and aqueous dissolution of the sulfide mineral oxidation products, including dissociated sulfuric acid and metals (e.g., Fe, Mn, Al), produces surface and groundwater degradation. Explanations of the chemical reactions by which acid mine drainage is produced from pyrite and other iron sulfide minerals are found in Singer and Stumm (1970), Kleinmann et al. (1981), Lovell (1983), Evangelou (1995), and Rose and Cravotta (1998). Additional references presenting data and discussion of factors related to pyrite oxidation rates include Emrich (1996), McKibben and Barnes (1986), Moses and Herman (1991), Watzlaf (1992), and Rimstidt and Newcomb (1993). These reactions also are presented and discussed in Section 2.0 of EPA's Coal Remining Best Management Practices Guidance Manual.

While pyrite is the most commonly reported producer of AMD, other mineral species including the sulfide mineral marcasite (FeS_2), and sulfate minerals jarosite ($KFe_3(SO_4)_2(OH)_6$) and alunite ($KAl_3(SO_4)_2(OH)_6$), are capable of producing acidic drainage at surface and underground mine sites. Sulfate minerals are generally secondary weathering products of pyrite oxidation. Nordstrom (1982) shows the sequence by which these minerals can form from pyrite. Many secondary sulfate minerals have been identified that are typically very soluble and transient in the humid eastern United States. These minerals form during dry periods and are flushed into the ground-water system during precipitation events. The sulfate minerals that contain iron, aluminum, or manganese are essentially stored acidity and will produce acid when dissolved in water. Sulfate minerals such as melanterite, pickeringite, and halotrichite occur commonly in Appalachian Basin coal-bearing rocks. Additional information about these sulfate minerals is found in Cravotta (1994), Lovell (1983), Rose and Cravotta (1998), and Brady et al. (1998).

Acid mine drainage is the most frequently described and most environmentally damaging type of coal mine drainage. However, other damaging types can occur due, principally, to geologic factors and influences from mining and reclamation practices. According to Rose and Cravotta (1998):

> "Coal Mine drainage ranges widely in composition, from acidic to alkaline, typically with elevated concentrations of sulfate (SO_4), iron (Fe), manganese (Mn), and aluminum (Al) as well as common elements such as calcium, sodium, potassium, and magnesium. The pH is most commonly either in the ranges 3 to 4.5 or 6 to 7, with fewer intermediate or extreme values... Acidic mine drainage (AMD) is formed by the oxidation of pyrite to release dissolved Fe^{2+}, SO_4^{2-} and H^+, followed by the further oxidation of the Fe^{2+} to Fe^{3+} and the precipitation of the iron as a hydroxide ("yellow boy") or similar substance, producing more H^+...In contrast, neutral or alkaline mine drainage (NAMD) has alkalinity that equals or exceeds acidity but can still have elevated concentrations of SO_4, Fe, Mn and other solutes. NAMD can originate as AMD that has been neutralized by reaction with carbonate minerals, such as calcite and dolomite, or can form from rock that contains little pyrite. Dissolution of carbonate minerals produces alkalinity, which promotes the removal of Fe, Al and other metal ions from solution, and neutralizes acidity. However, neutralization of AMD does not usually affect concentrations of SO_4."

The rate of AMD production and the concentrations of acidity, sulfate, iron, and other water quality parameters in mine drainage are dependent upon numerous physical, chemical, and biological factors. According to Rose and Cravotta (1998):

> "Many factors control the rate and extent of AMD formation in surface coal mines. More abundant pyrite in the overburden tends to increase the acidity of drainage, as does decreasing grain size of the pyrite. Iron-oxidizing bacteria and low pH values speed up the acid-forming reaction. Rates of acid formation tend to be slower if limestone or other neutralizers are present. Access of air containing the oxygen needed for pyrite oxidation is commonly the limiting factor in rate of acid generation. Both access of air and exposure of pyrite surfaces are promoted by breaking the pyrite-bearing rock. The oxygen can gain access either by molecular diffusion through the air-filled pore space in the spoil, or by flow of air which is driven through the pore space by temperature or pressure gradients..."

Numerous studies have evaluated the distribution of total sulfur contents and pyritic sulfur contents within coals and overburden strata. In some of these studies, investigations have examined the significance of pyrite morphology, especially the framboidal form with high surface area.

AMD discharges in Pennsylvania range in flow from seeps of less than 1 gallon per minute (gpm) to abandoned underground mine outfalls such as the Jeddo Tunnel near Hazleton, PA where a flow greater than 150,000 gpm (40,000 gpm average flow) has been measured. Table 2.0a presents typical and extreme examples of acidity, alkalinity, and related water quality parameters in coal mine drainage (from surface mines, underground mines, and coal refuse piles) and nearby well and spring samples. These water samples were compiled from data in Hornberger and Brady (1998) and Brady et al. (1998) to illustrate mine drainage quality variations in Pennsylvania. Similar variations in mine drainage quality exist in West Virginia, Ohio, and other states in the Appalachian Basin. Acidity and alkalinity concentrations greater than 100 mg/L are shown in bold in Table 2.0a.

Some of the most extreme concentrations of acidity, iron, and sulfate in Pennsylvania coal mine drainage, have been found at the Leechburg Mine refuse site in Armstrong County, and at surface mine sites in Centre, Clinton, Clarion, and Fayette Counties (Table 2.0a). Acidity concentrations of seeps from Lower Kittanning Coal refuse at the Leechburg site exceed 16,000 mg/L, while the sulfate concentration of one sample exceeds 18,000 mg/L. Schueck et al. (1996) reported on AMD abatement studies conducted at a backfilled surface mine site in Clinton County. A monitoring well that penetrated a pod of buried coal refuse produced a maximum acidity concentration of 23,900 mg/L prior to the implementation of the abatement measures.

Table 2.0a: High Alkalinity Examples in Pennsylvania Mine Discharges

Site Name	Stratigraphic Interval	pH	Alkalinity mg/L	Acidity mg/L	Fe mg/L	Mn mg/L	SO_4 mg/L	Flow gpm	Comments
Willow Tree	Waynesburg	7.8	**379.0**	0.0	0.12	0.04	165.0	1.0	Seep at deep mine, pre-mining
Susan Ann	Waynesburg	3.3	0.0	**1500.0**	324.40	89.70	2616.0	< 1.0	Seep near sealed deep mine entry
Bertovich	Sewickley	3.1	0.0	**378.0**	74.80	9.14	1098.0	2.0	Deep mine discharge
Smith	Redstone	7.7	**246.0**	0.0	1.47	0.27	122.0	0.0	Pit water at lowwall sump
Brown	Redstone	7.4	**626.0**	0.0	1.65	1.05	1440.0	no data	Spring near cropline
Trees Mills	Pittsburgh	2.5	0.0	**3616.0**	190.40	13.50	1497.8	13.0	Deep mine discharge
State Line	Upper & Lower Bakerstown	8.1	**210.0**	0.0	< 0.30	1.37	416.0	no data	Post-mining seep from backfilled spoil
Cover Hill	Lower Bakerstown	3.6	0.0	**168.1**	0.83	14.60	787.0	1.8	Discharge from abandoned pit below site
Hager	Brush Creek	6.8	**189.4**	no data	0.21	0.40	68.2	4.0	Logan spring
Fruithill	Upper & Lower Freeport	7.8	**238.0**	0.0	0.01	0.01	458.0	60.0	Deep mine discharge
Laurel Hill #1	U. Freept. to U. Kittng.	8.1	**484.0**	0.0	0.97	1.98	590.0	no data	Toe of spoil seep
Morrison	Upper Kittanning	7.0	**308.0**	0.0	0.63	3.49	327.0	< 1.0	Seep near collection ditch
Stuart	Upper Kittanning	2.8	0.0	**1290.0**	56.70	49.20	1467.0	no data	Seep, sandstone overburden
Clinger	Middle Kittanning	6.8	**190.0**	0.0	< 0.30	1.28	184.0	0.0	Pit water
Leechburg	Lower Kittanning	2.4	0.0	**16718.0**	> 300.0	19.30	18328.0	2.0	Seep from coal refuse disposal area
* Fran	Lower Kittanning	2.2	0.0	**23900.0**	5690.00	79.00	25110.0	0.0	Monitoring well in backfilled spoil
Swiscambria	Lower Kittanning	4.2	5.0	88.0	0.09	24.20	1070.0	no data	Seep, freshwater paleoenvironment
Albert #1	Lower Kittanning	3.1	0.0	**1335.0**	186.00	111.00	3288.0	55.0	Spoil discharge, brackish paleoenviron.
Snyder #1	Lower Kittanning	6.9	**114.0**	0.0	1.10	3.14	264.0	0.0	Pit water, marine paleoenvironment
Lawrence	Lower Kittanning	2.2	0.0	**5938.0**	2060.00	73.00	3600.0	0.0	Pit water, sandstone overburden
Graff Mine	L. Kittng. & Vanport Ls	7.8	**274.0**	no data	0.01	1.13	1645.0	10.0	Seep above road
Philipsburg	Clarion	2.7	0.0	**9732.0**	1959.80	205.30	4698.0	35.0	Spoil discharge
** Old 40	Clarion	2.2	0.0	**10000.0**	3200.00	260.00	14000.0	0.0	Monitoring well in backfilled spoil
Orcutt	Clarion	3.9	0.0	**5179.6**	2848.00	349.00	11120.0	0.0	Spoil water from piezometer

Site Name	Stratigraphic Interval	pH	Alkalinity mg/L	Acidity mg/L	Fe mg/L	Mn mg/L	SO_4 mg/L	Flow gpm	Comments
Cousins	Clarion	7.6	**130.0**	0.0	7.15	0.30	71.0	0.0	Pit water, glacial till influence
Zacherl	Clarion	2.3	0.0	**9870.0**	2860.00	136.60	7600.0	no data	Toe of spoil discharge
Horseshoe	Mercer	2.3	0.0	**1835.0**	194.00	27.00	2510.0	700.0	Abandoned deep mine discharge
Wadesville	Llewellyn	6.7	**414.0**	0.0	3.61	3.37	1038.0	no data	Minepool, Anthracite Region

Note: Extreme values (>100 mg/L) of alkalinity and acidity are highlighted for emphasis
* data from Schuek et al. (1996)
** data from Dugas et al. (1993)

Since the alkalinity-production process has a dramatically different set of controls, the resultant maximum alkalinity concentrations found in mine environments are typically one or two orders of magnitude less than the maximum acidity concentrations. Examples of relatively high alkalinity concentration in mine drainage, ground water, and surface water associated with Pennsylvania bituminous and anthracite coal mines are presented in Table 2.0a. The highest natural alkalinity concentration found in PA DEP mining permit file data (and reported in Table 2.0a) is 626 mg/L in a spring located near the cropline of the Redstone Coal in Fayette County. Thick sequences of carbonate strata, including the Redstone Limestone and the Fishpot Limestone underlie and overlie the Redstone Coal.

Carbonate minerals (e.g., calcite and dolomite) play an extremely important role in determining post-mining water chemistry. They neutralize acidic water created by pyrite oxidation, and there is evidence that they also inhibit pyrite oxidation (Hornberger et al., 1981; Williams et al., 1982; Perry and Brady, 1995). Brady et al. (1994) concluded that the presence of as little as 1 to 3 percent carbonate (on a mass-weighted basis) at a mine site can determine whether that mine produces alkaline or acid water. Although pyrite is clearly necessary to form acid mine drainage, the relationship between the amount of pyrite present and water-quality parameters (e.g., acidity) was only evident where carbonates were absent (Brady et al., 1994).

The paleoclimatic and paleoenvironmental influences on rock chemistry in the northern Appalachians resulted in the formation of coal overburden with greatly variable sulfur content (0 percent to >15 percent S) and calcareous mineral content (0 percent to >90 percent $CaCO_3$) as shown on figures of overburden drill hole data in Brady et al. (1998). The wide variations in rock chemistry contribute to the wide variations in water quality associated with surface coal mines. Figures 2.0a and 2.0b show the frequency distributions (i.e., range) of pH in mine discharges in the bituminous and anthracite coal regions of Pennsylvania. The origin and significance of this bimodal frequency distribution for mine drainage discharges are described in Brady et al. (1997, 1998) and Rose and Cravotta (1998). Brady et al. (1997) explained that although pyrite and carbonate minerals only comprise a few percent (or less) of the rock associated with coal, these acid-forming and acid-neutralizing minerals, respectively, are highly reactive and are mainly responsible for the bimodal distribution. Depending on the relative abundance of carbonates and pyrite, and the relative weathering rates, the pH will be driven toward one mode or the other.

Variations in the chemical composition of mine drainage discharges are principally related to geologic and hydrologic factors. The hydrologic factors that cause individual mine drainage discharges to vary in flow and concentrations of acidity, alkalinity, sulfates and metals (e.g., Fe, Mn, Al) throughout the water year are discussed in the following sections of this chapter.

Figure 2.0a: Distribution of pH in Bituminous Mine Drainage

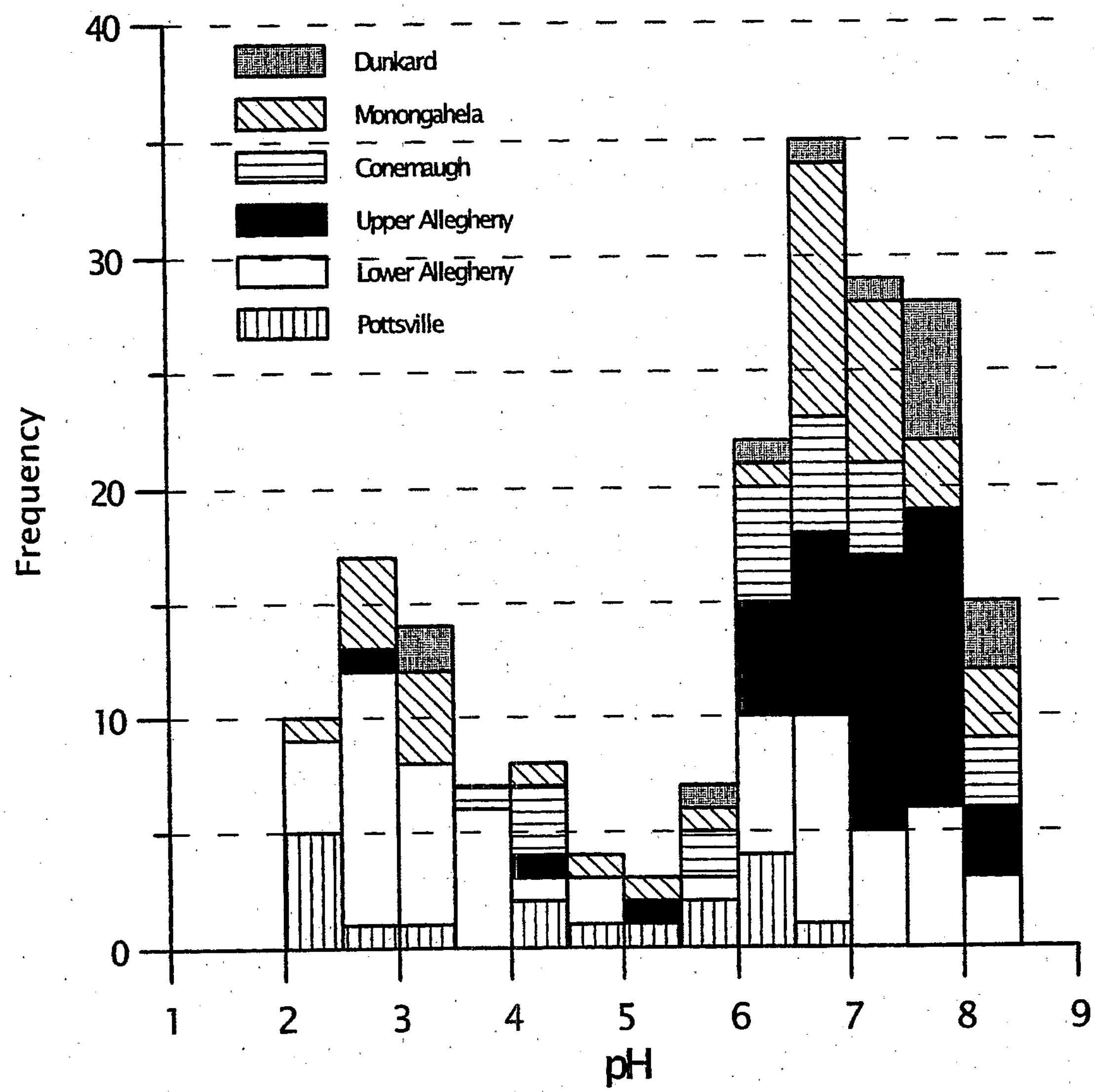
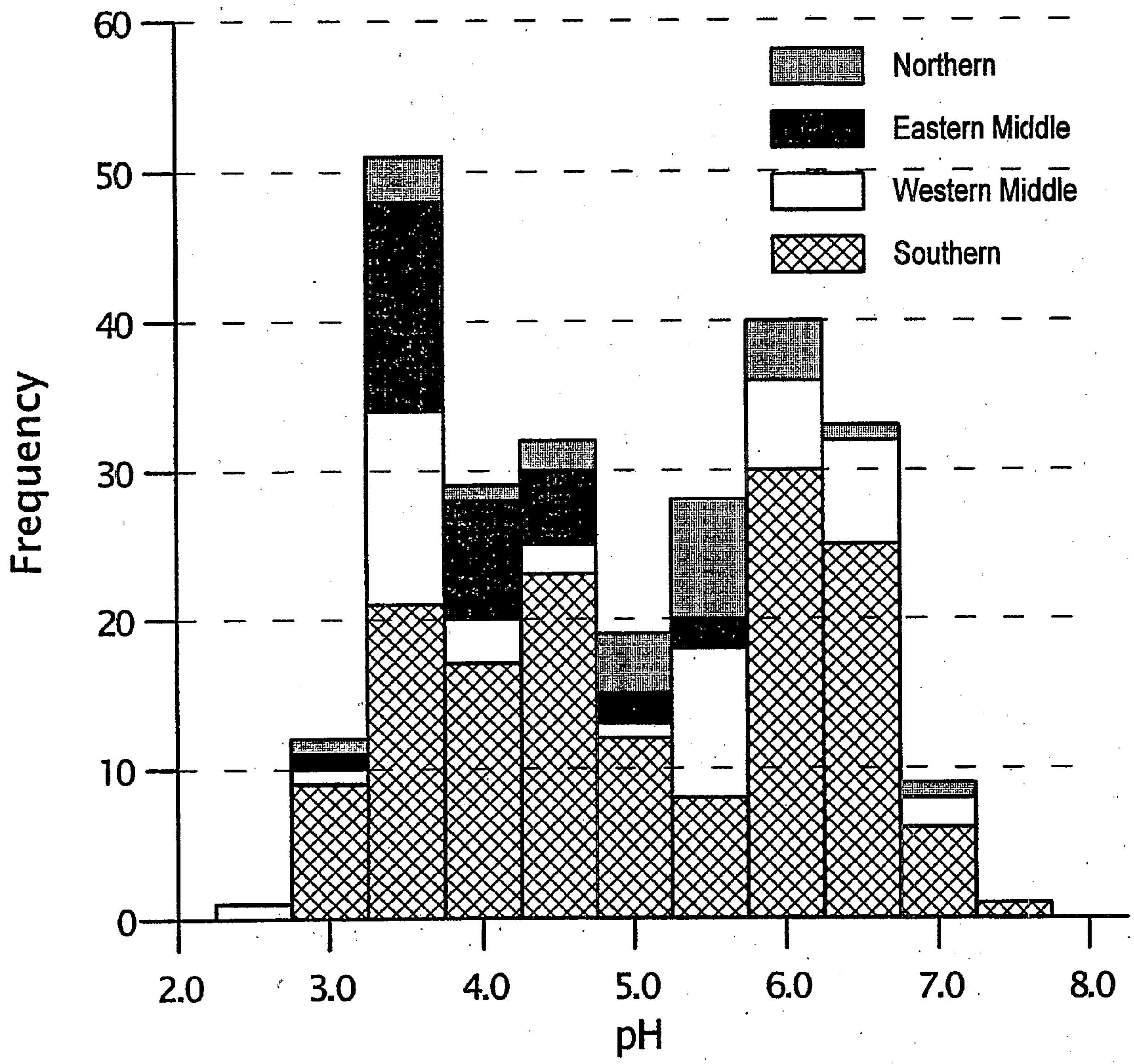


Figure 2.0b: Distribution of pH in Anthracite Mine Drainage

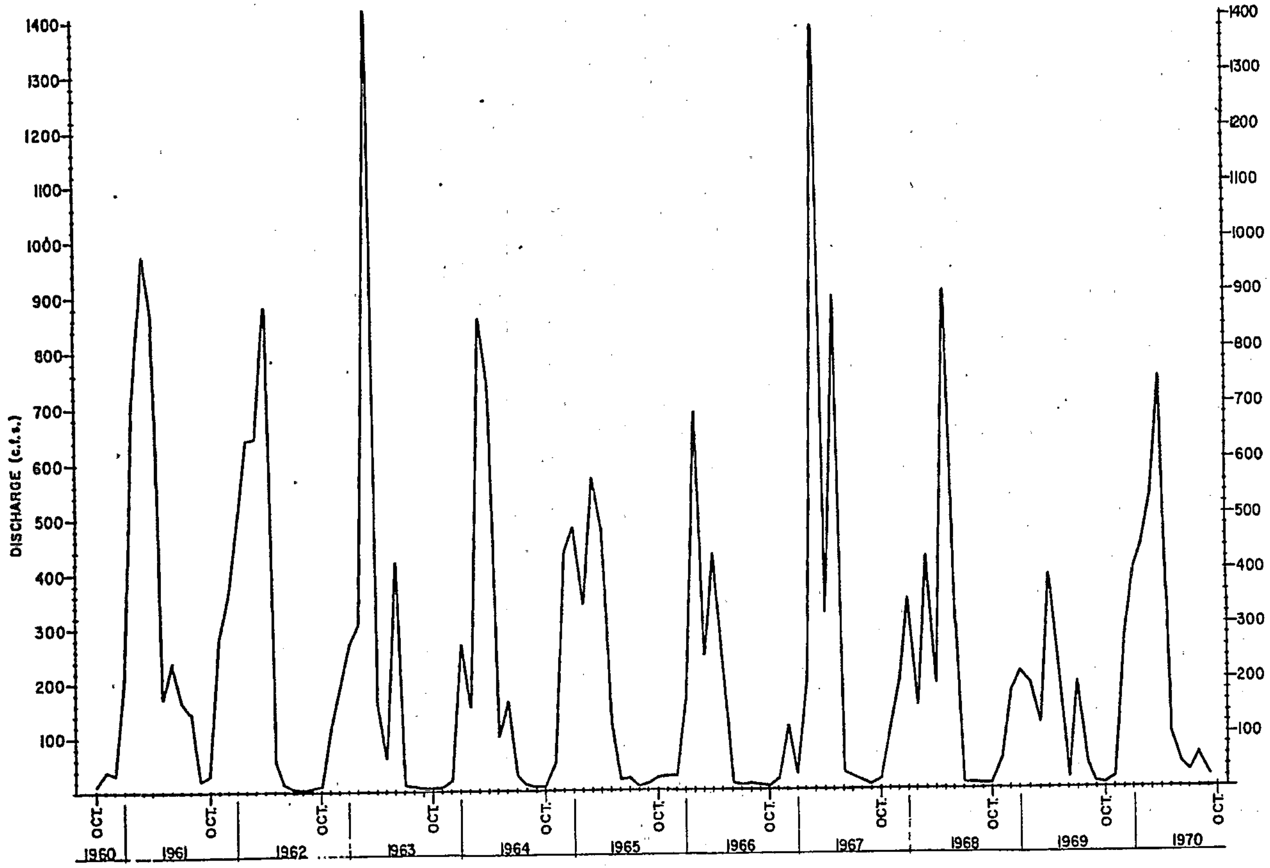


2.1 Impact of Stream Flow Variation on Water Quality Parameters

Annual variations in streamflow and surface water quality degraded by AMD discharges can be very significant as shown in Hornberger et al. (1981) for water quality network stations including small streams and large rivers in western Pennsylvania. These water quality network stations are closely monitored by PADEP. The streams are sampled several times yearly and analyzed for a wide array of water quality parameters, and usually are located in close proximity to U.S. Geological Survey (USGS) stream hydrograph stations for which extensive streamflow data area compiled and published. The data from the network stations is contained in the STORET database maintained by EPA.

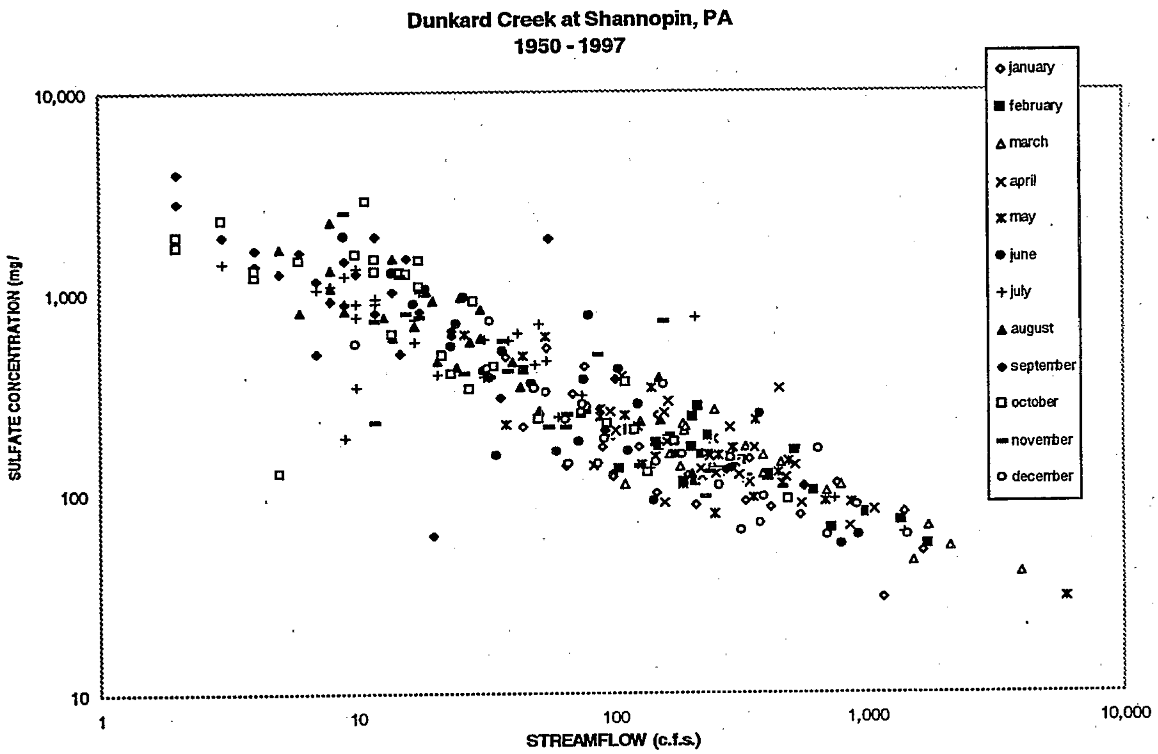
The water quality network station with the greatest range in streamflow and concentration of AMD related water quality parameters is the Dunkard Creek Station, in Greene County, Pennsylvania (Hornberger et al., 1981). This compilation includes greater than 150,000 lines of STORET data. Streamflow varied between 2 cubic feet per second (cfs) and 4,020 cfs in approximately 100 samples collected between 1950 and 1976, while the concentration of sulfates ranged from 40 to 4000 mg/L. The annual cycles of streamflow variations from October 1960 to September 1970 for Dunkard Creek are shown in Figure 2.1a, which was plotted by Hornberger et al. (1981) from monthly means of discharge data compiled by the U.S. Geological Survey.

Figure 2.1a Annual Variability in Streamflow at Dunkard Creek



In order to examine the relationship between variations in streamflow and corresponding variations in a reliable water quality indicator parameter, a logarithmic plot of sulfate concentration versus discharge was constructed using procedures described in Gunnerson (1967), Hem (1970), and Hornberger et al. (1981). The sulfate concentrations in Dunkard Creek tend to systematically decrease with increasing flow as shown by the approximately linear inverse relationship on Figure 2.1b. However, the relationship between streamflow and concentration may be more appropriately defined by a general elliptical progression of monthly flow and water quality relationships surrounding a least squares line fitted to the data points, similar to that found by Gunnerson (1967) and Hornberger et al. (1981). The tendency for high flow accompanied by low sulfate concentration in January, February, March, April, and May and low flow accompanied by high sulfate concentration in July, August, September, and October, and other flow-quality relationships throughout the water year may be observed in Figure 2.1b. Figure 2.1b includes almost 50 years of data (1950-1997) that show a stronger inverse linear relationship between sulfate concentration and streamflow than was shown in the first 26 years of data (Hornberger et al., 1981). The correlation coefficient (r) between sulfate concentration and streamflow data in Figure 2.1b is -0.887 (for logarithmically transformed data), which is statistically significant at the 1 percent level (N=307). The coefficient of determination [r^2] for this dataset is 0.787; therefore, 78.7 percent of the variations in sulfate concentration of the Dunkard Creek are accounted for by variations in streamflow. Similar patterns of variation in sulfate concentration and flow of a major AMD-impacted river were found (Hornberger et al., 1981) for the West Branch Susquehanna River at Renovo, Pennsylvania.

Figure 2.1b: Sulfate Concentration vs. Streamflow at Dunkard Creek



2.2 AMD Discharge Types and Behaviors

Discharges of acid mine drainage (AMD) can exhibit very different behavior depending upon the type of mine involved and its geologic characteristics. The hydrologic characteristics of a pre-existing AMD discharge can have important ramifications for documenting baseline pollution load – affecting the frequency and duration of sampling required to obtain a representative baseline. Braley (1951) was among the first to study the hydrology of AMD discharges. He noted that, much like a stream, flow rates vary dramatically in response to precipitation events and seasons, and that acid-loading rates are chiefly a function of flow. The greater the flow, the greater the load. Smith (1988), looking at long-term records of AMD discharges in Pennsylvania, classified discharges based on three fundamental behaviors: 1) High flow - low concentration / low flow - high concentration response, where the flow rate varies inversely with concentration; 2) Steady response where changes in flow rate and chemistry are minimal or damped; and 3) "Slug" response where large increases in discharge volumes are not accompanied by corresponding reductions in concentrations, resulting in large increases in pollution loading.

Figure 2.2a presents the discharge and acidity hydrograph of a mine discharge exhibiting the first (high flow - low concentration / low flow - high concentration) behavior. This discharge drains from a relatively small underground mine complex (Duffield, G.M., 1985). Typical for this type of discharge, the flow rate varies greatly and is subject to seasonal flow variations as well as individual precipitation events. Acidity concentrations vary inversely with the discharge rate, with the highest acidity occurring during the low-flow months of September, October, and November. The inverse log-linear relationship between discharge and acidity is shown in Figure 2.2b. Acidity steadily decreases with increasing flow, reflecting dilution of the mine drainage during periods of abundant ground-water recharge. Nonetheless, the pollution loading (i.e., the total acidity produced from the discharge in pounds per day) increases during high-flow events, as the decrease in acidity is not commensurate with a given increase in flow. In this sense, the discharge behaves very much like a stream and is subject to large increases in flow which dilute the concentration of dissolved chemical constituents. However, concentration decreases are not enough to offset flow increases. Pollution loading tends to parallel the flow rate but in a more subdued manner. The majority of pre-existing AMD discharges in Pennsylvania exhibit this type of behavior. It is most common for surface mine discharges and discharges from small to medium size underground mines where the capacity for ground water storage is relatively small and ground water flow paths are short.

Some discharges, particularly large-volume discharges from extensive underground mine complexes, show comparatively little fluctuation in discharge rate and only minor variation in chemical quality. Figure 2.2c presents such an example from a Schuylkill County, Pennsylvania, anthracite underground mine. In this case, the exceptionally large recharge area and volume of water in the mine pool, and the stratification of water quality within the mine pool, are causing a steady-response behavior of the discharge. Short-term fluctuations in flow and quality are subdued, because of the large amount of stored ground water acting as a reservoir and dampening fluctuations due to individual recharge events.

Figure 2.2a: Acidity and Streamflow of Arnot Mine Discharge

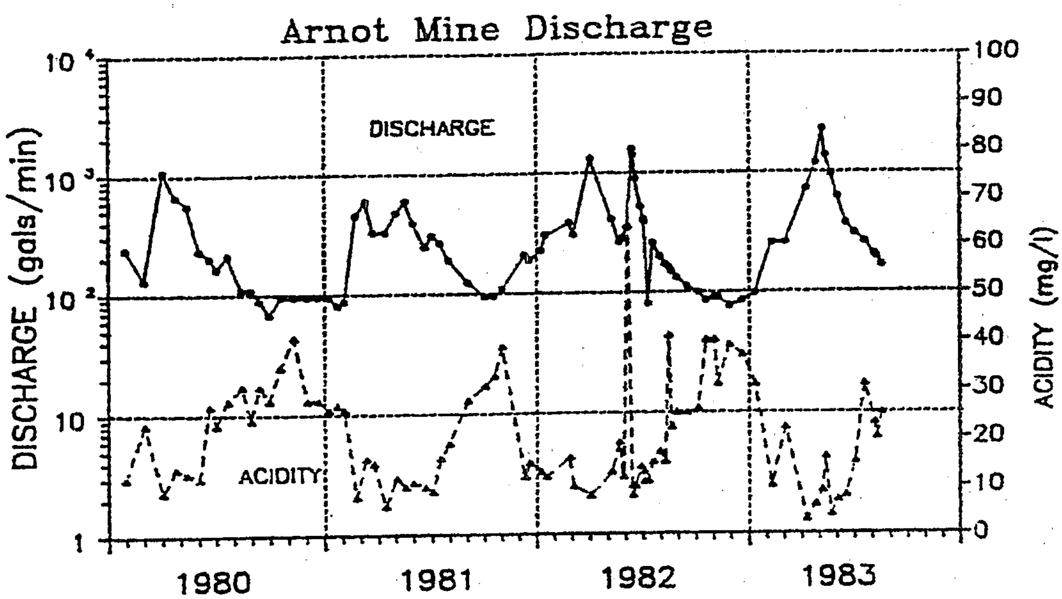
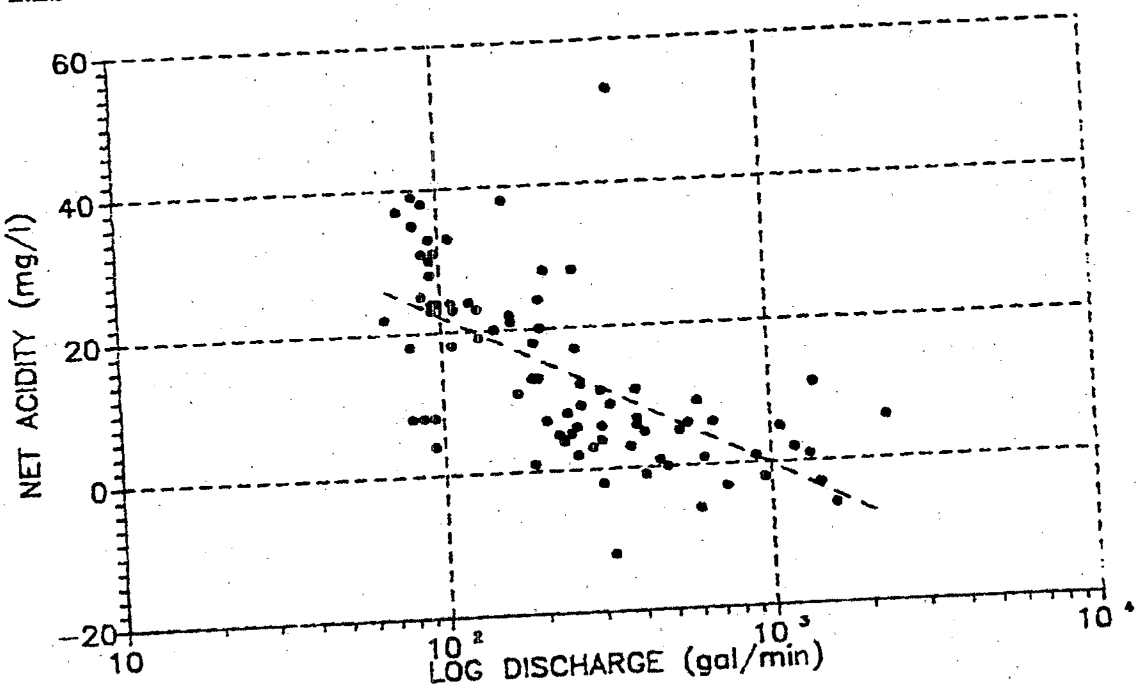


Figure 2.2b: Inverse Loglinear Relationship between Acidity and Streamflow



Occasionally, AMD discharges are subject to extreme variations in flow rates with little change in water quality. Figure 2.2d presents flow and acidity exhibiting "slug" behavior in a discharge from a coal refuse pile. Flow rates vary dramatically in response to recharge events (from less than 3 to 470 gpm). Concomitantly, acidity concentrations change very little and result in large, rapid variations in acid loading. This discharge behavior results where conditions favor the accumulation of water-soluble, acid-bearing shales in the unsaturated zone. During recharge events, infiltrating water permits rapid dissolution of salts producing additional acidity in the discharge, rather than causing a dilution effect. The longer the time period between recharge events, the more time is available for the build up of acid-bearing salts in the unsaturated zone. Coal refuse piles, and surface mines with very high sulfur spoil in the unsaturated zone and limited ground-water storage capacity, provide the most favorable environment for this discharge behavior. In the most severe cases, increases in flow can be accompanied by increased concentrations of acidity or metals, resulting in extreme increases in loading rates. When this

Figure 2.2c: Streamflow and Acidity in Schuylkill County

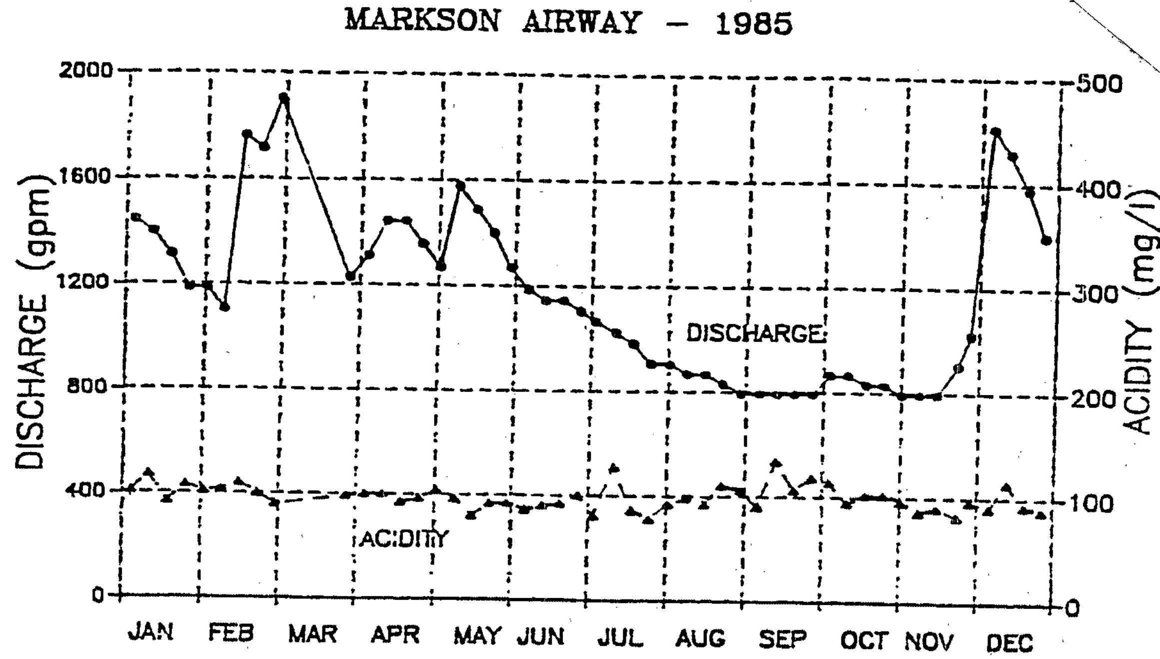
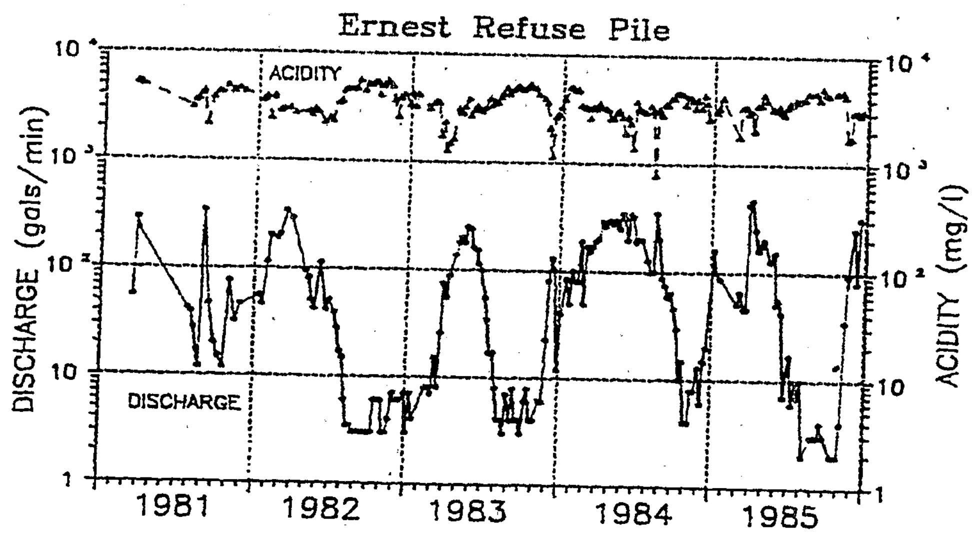


Figure 2.2d: Streamflow and Acidity in Coal Refuse Pile



phenomenon occurs on a large scale, potentially disastrous increases in acid loading can adversely affect downstream water uses and aquatic life.

The Arnot, Markson, and Ernest mine drainage discharges described in the preceding paragraphs were originally studied and graphically presented in Smith (1988) and Hornberger et al. (1990). These three mine drainage discharges are also the subject of three of the eight water quality reports completed by Griffiths (1987, 1988) as part of the cooperative EPA/PADEP remining project and included in the abridged volume by EPA (2001, EPA-821-B-01-014).

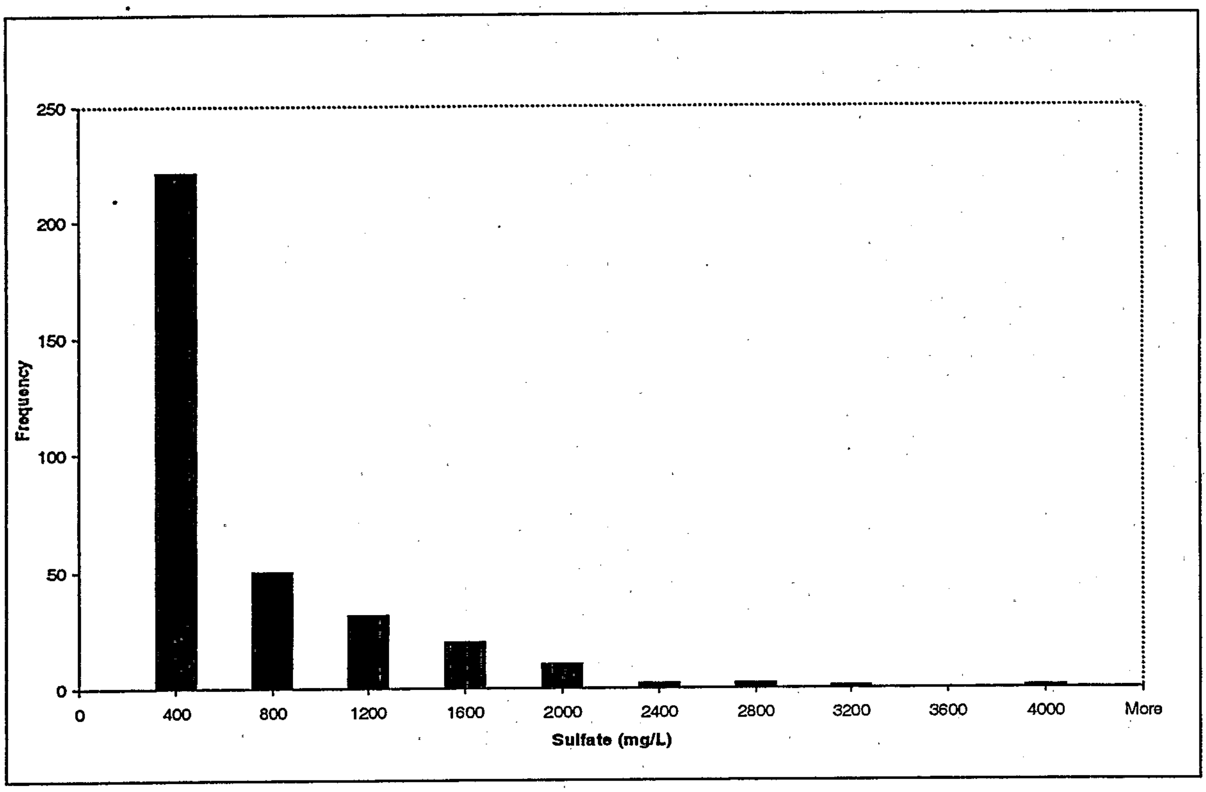
For remining operations that will reaffect a pre-existing pollutional discharge, knowledge of discharge behavior is critical to the establishment of a representative baseline. All three discharge types exhibit some seasonal behavior, with highest flows during seasonal high ground-water conditions and the lowest flows and loadings during low ground-water conditions. For most of Appalachia, high ground-water conditions occur during late winter or spring. Low ground-water conditions occur during late summer and early fall. The baseline sampling period must cover the full range of seasonal conditions. Exactly when these extremes will occur is unpredictable, as storm events may occur over relatively short time intervals. Accordingly, to properly characterize an AMD discharge, it is usually necessary to monitor the discharge over at least an entire water year with a sufficiently narrow sampling interval to capture short-term extreme events. Slug-response discharges may require more frequent sampling due to their flashy hydrologic response with large variations in pollution load over short time intervals. Conversely, less frequent baseline sampling may be adequate for damped-response discharges.

Because the baseline is based on loading rates, accurate flow measurements are as important as contaminant concentration measurements. Previous studies by Smith (1988), Hornberger et al. (1990), and Hawkins (1994) have emphasized the strong relationship between flow rate and contaminant load. Hawkins (1994) analyzed pre- and post-remining hydrologic data from 24 remining sites in Pennsylvania and noted that flow was the dominant factor in changes in post-mining pollution loads. Most remining operations that reduced baseline pollution load did so by reducing the flow of the pre-existing discharge. In view of this, Smith (1988) points out that proper flow measurement is of overriding importance in determining the baseline pollution load.

2.3 Distributional Properties of AMD Discharges

Water quality parameters of many AMD discharges and AMD impacted streams are not normally distributed. In most cases these frequency distributions are highly skewed because there are many samples with relatively low concentrations and a few samples with very high concentrations due to low-flow drought conditions or slugs of pollution in response to major storm events. Plotting these data on a logarithmatic scale (as shown on Figure 2.1b), or logarithmically transforming the data produces a much closer approximation of the normal frequency distribution.

Figure 2.3a: Frequency Distribution of Sulfate at Dunkard Creek (mg/L)



Numerous variables with continuous data on the interval or ratio level of information exhibit log normal behavior in the natural environment (Aitchison and Brown, 1973; Krumbein and Graybill, 1965; Griffiths, 1967), and logarithms are frequently used in the analysis and graphical expression of water quality data (Gunnerson, 1967; Hem, 1970). The log normal distribution is also very common in previous EPA work with wastewater discharges. Figure 2.3a shows the skewed frequency distribution for the sulfate data for the Dunkard Creek dataset used in Figure 2.1b.

Examples of the distributional properties of data from AMD discharges at remining sites in Pennsylvania are shown in Figures 2.3b to 2.3f from the EPA publication Statistical Analysis of Abandoned Mine Drainage in the Assessment of Pollutant Load (EPA-821-B-01-014), which is a companion volume to this report. The figures show frequency distributions of data using stem-and-leaf diagrams. For additional information on stem-and-leaf diagrams, see Hoaglin et al. 1983.

Figure 2.3b shows a nearly normal frequency distribution of pH of the Arnot 003 discharge (N=82). An example of a highly skewed frequency distribution is given in Figure 2.3c for flow of the Clarion discharge. Following logarithmic transformation, the frequency distribution becomes more symmetrical, approaching normality, as seen in Figure 2.3d. However, some caution must be exercised in applying log transformations to data sets because overcorrection may occur. Such overcorrection is seen in the irregular frequency distribution of acidity concentration in the Clarion discharge. In Figure 2.3e, the untransformed data are somewhat positively skewed. Following transformation, these data become highly negatively skewed (Figure 2.3f).

Figure 2.3b: Stem-and-leaf Diagram of pH (Arnot 003)

N = 82

Leaf Unit = 0.010

1	30	4
2	30	7
5	31	134
17	31	555667888999
36	32	0011111112234444444
(15)	32	555666777789999
31	33	001111222234
19	33	556666778
10	34	1122
6	34	679
3	35	0
2	35	7
1	36	
1	36	
1	37	0

Figure 2.3c: Stem-and-leaf Diagram of Discharge

N = 77

Leaf Unit = 1.0

(53)	0	000000000011112223333333444445555555556666667788999999
24	1	00222244444
13	2	001288
7	3	066
4	4	0
3	5	0
2	6	
2	7	
2	8	3
1	9	
1	10	
1	11	
1	12	
1	13	
1	14	
1	15	
1	16	
1	17	2

Figure 2.3d: Stem-and-leaf Diagram of Log Discharge

N = 75

Leaf Unit = 0.10

4	-1	3000
7	-0	766
11	-0	4330
19	0	01124444
(35)	0	55555566666777777777888888999999
21	1	000111111333344
6	1	55679
1	2	2

Figure 2.3e: Stem-and-leaf Diagram of Acidity

N = 96

Leaf Unit = 10

8	0	00114568
21	1	0234466667788
30	2	033356899
38	3	01245778
(14)	4	01114456788899
44	5	0112778899
34	6	03344678
26	7	014499
20	8	15
18	9	1455568899
8	10	38
6	11	09
4	12	04
2	13	8
1	14	
1	15	4

Figure 2.3f: Stem-and-leaf Diagram of Log Acidity

N= 97

Leaf Unit = 0.10

1	-1	0
1	-0	
2	-0	0
3	0	3
3	0	
5	1	22
9	1	6679
31	2	0011122222222333444444
(58)	2	5555555566666666666667777777777777888888888889999999999999
8	3	00000011

References

Aitchison, J. and J.A.C. Brown, 1973. The Lognormal Distribution. Cambridge University Press: London. 176 p.

Barnes, H.L. and S.B. Romberger, 1968. Chemical Aspects of Acid Mine Drainage. Journal of the Water Pollution Control Federation, Vol. 40, No. 3, pp. 371 – 384.

Brady, K.B.C., E.F. Perry, R.L. Beam, D.C. Bisko, M.D. Gardner, and J.M. Tarantino, 1994. Evaluation of Acid-base Accounting to Predict the Quality of Drainage at Surface Coal Mines in Pennsylvania, U.S.A. Presented at the International Land Reclamation and Mine Drainage Conference on the Abatement of Acidic Drainage, Pittsburg, PA, April 24-29, 1994.

Brady, K.B.C., A.W. Rose, C.A. Cravotta, III, and W.W. Hellier, 1997. Bimodal Distribution of pH in Coal Mine Drainage. Abstracts with Programs, 27th Northeast Section Meeting, Geological Society of America, Vol. 24, No. 3.

Brady, K.B.C., R.J. Hornberger, and G. Fleeger, 1998. Influence of Geology on Postmining Water Quality: Northern Appalachian Basin. Chapter 8 in Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Edited by K.B.C. Brady, M.W. Smith, and J. Schueck. Department of Environmental Protection: Harrisburg, PA. pp. 8-1 to 8-92.

Braley, S.A., 1954. Summary Report on Commonwealth of Pennsylvania, Department of Health Industrial Fellowship (Nos. 1,2,3,4,5,6, and 7). Fellowship 326B, Mellon Institute of Industrial Research.

Cravotta, C.A., III, 1994. Chapter 23: Secondary Iron-sulfate Minerals as Sources of Sulfate and Acidity – the Geochemical Evolution of Acidic Ground Water at a Reclaimed Surface Coal Mine in Pennsylvania. US Geological Survey: Lemoyne, PA.

Duffield, G.M., 1985. Intervention Analysis Applied to the Quantity an Quality of Drainage from an Abandoned Underground Coal Mine in North-Central Pennslvania. Masters of Science Thesis, Pennsylvania State University, Department of Geology.

Dugas, D.L., C.A. Cravotta, III, and D.A. Saad, 1993. Water Quality Data for Two Surface Coal Mines Reclaimed with Alkaline Waste or Urban Sewage Sludge, Clarion County, Pennsylvania, May 1983 through November 1989. U.S.Geological Survey Open File Report 93 – 115: Lemoyne, PA.

Emrich, G.H., 1966. Tests for Evaluating the Quality of Mine Drainage Characteristics of Coal Seams. Pennsylvania Department of Health, Division of Sanitary Engineering, Mine Drainage Technical Bulletin No. 2., pp. V-6 thru V-14.

Evangelou, V.P., 1995. Pyrite Oxidation and its Control. CRC Press: New York, NY. 293 p.

Griffiths, J. C., 1967. Scientific Method in Analysis of Sediments. McGraw Hill Book Co.: New York, NY. 508 p.

Griffiths, J.C., (No date). Report No. 1: Reconnaissance into the use of MINITAB using Hamilton 01, 08 Data. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987a. Report No. 2: Time Series Analysis of Data from Hamilton 08. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987b. Report No. 3: Analysis of Data from Arnot 001, 003, 004. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987c. Report No. 4: Analysis of Data from the Clarion Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987d. Report No. 5: Analysis of Data from Ernest Refuse Pile, Indiana County, Pennsylvania. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987e. Report No. 6: Analysis of Data from the Fisher Deep Mine. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988a. Report No. 7: Analysis of Data from the Markson Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988b. Report No. 8: Synopsis of Reports 1 - 7. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Gunnerson, C.G., 1967. Streamflow and Quality in the Columbia River Basin. Journal of the Sanitary Engineering Division, Proceedings of the American Society Civil Engineers, Vol. 93, No. SA6, pp. 1 – 16.

Hawkins, J.W., 1994. Statistical Characteristics of Coal Mine Discharges on Western Pennsylvania Remining Sites. Water Resources Bulletin, Vol. 30, No. 5, pp. 861 – 869.

Hem, J.D., 1970. Study and Interpretation of the Chemical Characteristics of Natural Water. US Geological Survey Water Supply Paper No. 1473, 2nd ed. U.S. Government Printing Office: Washington, DC. 363 p.

Hoaglin, D.C., F. Mosfeller, and J.W. Tukey, 1983. Understanding robust and exploratory data analysis. John Wiley & Sons: New York, NY. pp. 7 – 32.

Hornberger, R.J., R.R. Parizek, and E.G. Williams, 1981. Delineation of Acid Mine Drainage Potential of Coal Bearing Strata of the Pottsville and Allegheny Groups in Western Pennsylvania. Final Report, Department. of Interior, Office of Water Research and Technology Project B-097-PA. Pennsylvania State University: University Park, PA.

Hornberger, R.J., M.W. Smith, A.E. Friedrich, and H.L. Lovell, 1990. Acid Mine Drainage from Active and Abandoned Coal Mines in Pennsylvania. Chapter 32 in Water Resources in Pennsylvania: Availability, Quality and Management. Edited by S.K. Majumdar, E.W. Miller and R.R. Parizek. The Pennsylvania Academy of Science: Easton, PA. pp. 432 – 451.

Hornberger, R.J., and K.B.C. Brady, 1998. Kinetic (Leaching) Tests for the Prediction of Mine Drainage Quality. Chapter 7 in Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Edited by K.B.C. Brady, M.W. Smith and J. Schueck. Department of Environmental Protection: Harrisburg, PA. pp. 7-1 to 7-54.

Kleinmann, R.L.P., D.A. Crerar, and R.R. Pacelli, 1981. Biogeochemistry of Acid Mine Drainage and a Method to Control Acid Formation. Mining Engineering, pp. 300 – 305, March 1981.

Krumbein, W.C., and F.A. Graybill, 1965. An Introduction to Statistical Models in Geology. McGraw Hill Book Company, Inc: New York, NY. 475 p.

Lovell, H.L., 1983. Coal Mine Drainage in the United States – An Overview. Water Science and Technology, Vol. 15, pp. 1 – 25.

McKibben, M.A. and H.L. Barnes, 1986. Oxidation of Pyrite in Low Temperature acidic Solutions – Rate Laws & Surface Textures. Geochimica et Cosmochimica Acta, Vol. 50, pp. 1509 – 1520.

Moses, C.O. and J.S. Herman, 1991. Pyrite Oxidation at Circumneutral pH. Geochimica et Cosmochimica Acta, Vol. 55, pp. 471 – 482.

Nordstrom, D.K., 1982. Aqueous Pyrite Oxidation and the Consequent Formation of Secondary Iron Minerals. Acid Sulfate Weathering, Special Publication No. 10, Soil Science Society of America, pp. 37 – 56.

Perry, E.F. and K.B.C. Brady, 1995. Influence of Neutralization Potential on Surface Mine Drainage Quality in Pennsylvania. In: Proceedings 16th Annual West Virginia Surface Mine Drainage Task Force Symposium, Morgantown, WV, April 4-5, 1995.

Rimstidt, J.D. and W.D. Newcomb, 1993. Measurement and Analysis of Rate Data – The Rate of Reaction of Ferric Iron with Pyrite. Geochimica et Cosmochimica Acta, Vol. 57, pp. 1919 – 1934.

Rose, A.W. and C.A. Cravotta, III, 1998. Geochemistry of Coal Mine Drainage. Chapter 1 in Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Edited by: K.B.C. Brady, M.W. Smith and J. Schueck Department of Environmental Protection: Harrisburg, PA. pp. 1-1 to 1-22.

Schueck, J., M. DiMatteo, B. Scheetz, and M. Silsbee, 1996. Water Quality Improvements Resulting from FBC Ash Grouting of Buried Piles of Pyritic Materials on a Surface Coal Mine. Proceedings of the 13th Annual Meeting – American Society for Surface Mining and Reclamation, Knoxville, TN, May 19-24, 1996.

Singer, P.C., and W. Stumm, 1970. Acidic Mine Drainage: The Rate Determining Step. Science, Vol. 167, pp. 1121 – 1123.

Smith, M.W., 1988. Establishing Baseline Pollution Load from Pre-existing Pollutional Discharges for Remining in Pennsylvania. Paper presented at the 1988 Mine Drainage and Surface Mine Reclamation Conference, Pittsburg, PA, April 17-22, pp. 311 – 318.

U.S. EPA, 2001. Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits. Prepared for US Environmental Protection Agency, Office of Water by Pennsylvania Department of Environmental Protection and DynCorp I&ET. EPA-821-B-01-014.

Watzlaf, G.R., 1992. Pyrite Oxidation in Saturated and Unsaturated Coal Waste. In: Proceedings, 1991 National Meeting of the American Society for Surface Mining and Reclamation, Duluth, MN. pp. 191 – 205, June 14-18, 1992.

Williams, E.G., A.W. Rose, R.R. Parizek, and S.A. Waters, 1982. Factors Controlling the Generation of Acid Mine Drainage. Final Report on U.S. Bureau of Mines Research Grant No. G5105086. Pennsylvania State University: University Park, PA.

Section 3.0 Statistical Methods for Establishing Baseline Conditions and Setting Discharge Limits at Remining Sites

3.1 Objectives & Evaluation of Statistical Methods

The Rahall amendment, CWA Section 301(p), states in part:

> *(2) LIMITATIONS. - The Administrator or the State may only issue a permit pursuant to paragraph (1) if the applicant demonstrates to the satisfaction of the Administrator or the State, as the case may be, that the coal remining operation will result in the potential for improved water quality from the remining operation but in no event shall such a permit allow the pH level of any discharge, and in no event shall such a permit allow the discharges of iron and manganese, to exceed the levels being discharged from the remined area before the coal remining operation begins."*

EPA has promulgated the Coal Remining Subcategory (40 CFR Part 434.70) consistent with the requirements and intent of the Rahall amendment. The regulations for the Coal Remining Subcategory establish a standardized statistical procedure for determining baseline pollutant loadings and pollutant loadings during remining for net acidity, solids, iron, and manganese in pre-existing discharges. These statistical procedures are codified in Appendix B to Part 434 and are intended to identify increases (during remining) of discharge pollutant loadings above the baseline levels.

EPA has interpreted "levels" to mean the entire probability distribution of loadings, including the average, the median, and the extremes. It follows that if P percent of loadings did not exceed some number Lp during baseline, then no more than P percent should exceed Lp during and after remining. For example, if during the baseline period, 95 percent of iron loadings are ≤ 8.1

lbs/day and 50 percent are ≤ 0.3 lbs/day, then during and after remining the same relationships should hold true.

The objective of Section 3 is to provide statistical procedures for deciding when the pollutant loadings in a discharge exceed the levels of baseline. These procedures are intended to provide a good chance of detecting a substantial, continuing state of exceedance, while reducing the likelihood of a "false alarm." The procedures (or the numbers calculated from them) are also referred to here as "triggers."

In developing these procedures, EPA considered the statistical distribution and characteristics of discharge loadings data from pre-existing discharges, the suitability of parametric and nonparametric statistical procedures for such data, the number of samples required for these procedures to perform adequately and reliably, and the balance between false positive and false negative decision error rates. EPA also considered the cost involved with sample collection as well as delays in permit approval during the establishment of baseline, and considered the potential that increased sampling could discourage remining. In order to sufficiently characterize pollutant levels during baseline determination and during each annual monitoring period, Appendix B to Part 434 requires that the results of a minimum of one sample be obtained per month for a period of 12 months.

The procedures described below will provide limits for both single observations and annual averages. This is intended to provide checks on both the average and extreme values. There is a need to take into account the number of observations used to determine compliance when setting a limit or when otherwise determining compliance with baseline. For example, the collection of a greater number of samples from a discharge will reduce the variability of the average level (provided that samples are distributed randomly or regularly over the sampling year). Accordingly, the statistical procedures described here take into account the amount of data in an appropriate fashion.

Use of a statistical decision procedure should result in suitable error rates. Technically these are usually referred to as the rate alpha (α) for Type I errors and the rate beta (β) for Type II errors. The error of concluding that an exceedance has occurred when the discharge is exactly matching the baseline condition is intended to happen with probability α. Alpha can be characterized as the maximum "false alarm rate." When the discharge level is substantially less than baseline, the probability of making this error is expected to be very low. The error of concluding that no exceedance has occurred, when the discharge has in fact exceeded baseline levels, is intended to happen with probability β. Power (π) equals $1-\beta$. Power can be defined as the probability that a statistical decision procedure will declare that remining loadings exceed baseline loadings when there really has been an increase, or as the rate of giving correct alarms.

When many decisions will be made, the overall error rate is a concern. For example, the single-observation triggers described below will be applied every month during remining; the annual triggers will be applied every year. In evaluating statistical methods, EPA considered the overall or cumulative decision error rates during a five-year period of compliance monitoring.

The degree of serial correlation of the data will influence the decision error rates of statistical procedures. There is significant, positive serial correlation of flow, concentration, and loading in mine discharges over periods of 1 to 4 weeks, that is, sequential samples are correlated with each other (U.S.E.P.A., 2001a, 2001b). Also, estimates of the variance, used in parametric statistical procedures, are inaccurate unless adjusted for autocorrelation. (Loftis and Ward, 1980; U.S.E.P.A., 1993). Such adjustments require an estimate of the autocorrelation coefficient. However, one cannot reliably estimate site-specific autocorrelation from small samples (e.g., n=12). Using long-term datasets for pre-existing discharges at abandoned mines and at remining sites, EPA estimated the first-order serial correlations (at a monthly interval) for flow and for iron, manganese, and acidity loadings. The estimates fell mostly in the range 0.35 to 0.65, with a central tendency just below 0.50 (U.S.E.P.A., 2001b).

EPA evaluated parametric and nonparametric statistical procedures for characterizing the baseline level and determining compliance with the baseline level (U.S.E.P.A., 2001c). For the evaluation, EPA simulated discharge loadings data. These data had realistic statistical properties, resembling actual discharge loadings in terms of distribution and serial correlation (U.S.E.P.A., 2001b, 2001c, 2001d). The data simulated a 1- year (12 month) baseline period followed by a 5-year (60-month) remining period, with loadings measured once every month (also weekly, when the procedure required a period of accelerated monitoring). The evaluation examined the ability of a number of statistical procedures to react to various degrees of decrease and increase in loadings after baseline. The parametric procedures employed appropriate adjustments to the estimated variance to account for first-order serial correlation (assumed to be 0.5). The evaluation assumed that a minimum of 12 measurements of pollutant loads were made every year, once each month.

The ideal statistical procedure would always declare "not larger" when remining pollutant loadings are less than or equal to baseline loadings, and would always signal "larger" when remining loadings exceeded baseline. No such ideal procedure exists. Instead, the rate of signalling "larger" will increase as the average difference between baseline and remining loadings increases in magnitude. Statistical triggers may be "tuned", by choosing their numerical constants, so that a compromise is achieved between false alarms, that is, signalling "larger" when remining loadings are not larger than baseline loadings, and correct alarms, when remining loadings truly are greater.

The evaluations led to a choice of procedures and of numerical constants that achieve a reasonable balance between false alarms and correct alarms. This reasonable balance was considered to be achieved when a trigger produced the following results:

> (a) when there was no change in loadings from the baseline to the remining time period, the "false alarm rate" (type-I error rate) was not larger than that for the triggers used by the Commonwealth of Pennsylvania. Pennsylvania's trigger was used as a benchmark because of the demonstrated success of this approach (Hawkins 1994).

(b) when the mean pollutant load increased by one standard deviation after the baseline period, statistical power (probability of detecting the increase) was at least 0.75.

(c) when there was a decrease of 0.5 standard deviations in the mean loading after the baseline period, the"false alarm rate" was smaller than 5%.

(d) when the mean loading increased by 1 to 2 standard deviations after the baseline period, the "correct alarm rate" (power) was maximized (compared with other procedures).

Details of EPA's evaluation and comparisons of statistical procedures are provided in a separate document (U.S.E.P.A., 2001c). EPA reached the following conclusions about the statistical triggers based on these evaluations.

(1) The magnitude of serial correlation has a substantial effect on power. Statistical triggers that have reasonable power when there is no serial correlation could be unreasonable when there is substantial serial correlation, because they could then have very high rates of type I errors (false alarms). It was necessary to select numeric constants for the statistical triggers that are appropriate to data having autocorrelation. For evaluating and comparing statistical methods and triggers, EPA assumed a first-order autocorrelation coefficient of 0.5.

(2) To avoid false alarms, EPA determined that sequential exceedances of the Single Observation Trigger and accelerated monitoring were necessary. This method has long been used successfully in Pennsylvania's Remining Program. Specifically, the Single Observations Trigger requires the following: "If two successive monthly monitoring observations both exceed L, immediately begin weekly monitoring for four weeks (four weekly samples). If three or fewer of the weekly observations exceed L, resume monthly monitoring. If all four weekly observations exceed L, the baseline pollution loading has been exceeded."

(3) In Method 2, the Annual Comparison was set such that tables for the 99.9% level (alpha = 0.001) rather than the 95% level (alpha = 0.05) are to be used for the Wilcoxon-Mann-Whitney Test. When Type I Error rates of alpha = 0.05 or 0.01, the Wilcoxon-Mann-Whitney test in Method 2 had a high rate of declaring loadings to be larger than baseline when if fact, they were not larger.

(4) Method 1 and Method 2 were both designed as nonparametric rather than parametric procedures, with power comparable to that of a parametric procedure. Unlike a parametric method which would require log-transformation, the nonparametric methods accommodate zero flows (which may occur during remining) and negatively valued data (which may occur for net acidity) without requiring additional or complex modifications.

(5) EPA believes that the error rates and power of these triggers are acceptable in practice because best management practices (BMPs) reduced discharge loadings substantially. Hawkins (1994) reviewed the application of these triggers to remining operations in Pennsylvania, and concluded that the rates of triggering were low because remining usually reduced loadings substantially. EPA's BMP guidance manual includes an extensive analysis of remining discharges that supports this conclusion (EPA, 2001a). EPA concluded that the statistical triggers that Pennsylvania uses in its remining program are acceptable and effective and has used them as the basis for Method 1 with minor modifications to meet the criteria in (a) to (d). Method 1 herein follows the Pennsylvania triggers exactly except that a constant (1.815 = 1.96 * 1.25 / 1.35) is used in the formula for the Annual Procedure (see McGill, Tukey, and Larsen, 1978). Pennsylvania uses a more stringent number (1.58 = 1.7 * 1.25 / 1.35).

(6) The evaluation of the false alarm rate applies to a worst-case situation. The rate of declaring loadings to be larger than baseline when they are not is overstated by the evaluations (U.S.E.P.A., 2001c). It is evaluated in terms of the percentage of mines that would experience at least one determination that loadings exceed the baseline level over a period of five years (60 months), when in fact there has been no change from baseline. In practice, the area contributing to a discharge should be remined and regraded in less time, after which the discharge flow and loading will be substantially reduced. Thus the time period during which pollutant loadings are monitored for each discharge will usually be

shorter than five years. This in turn will mean lower percentages of false positives and false negatives than reported in Table 3.1a.

The power of statistical triggers for the final regulation is shown in Table 3.1a. The results show that Method 1 and Method 2 have comparable power to detect large increases (see columns '1σ' and '+2σ'). The main difference stems from the Monthly (Single Observation Limit) Test, which has higher false alarm percentages (see columns labeled '-0.5σ' and '0') when Method 1 is used.[1] Note that the Annual Comparison used without the Single Observation Limit Test would not have a high rate of detecting an increase of one standard deviation above baseline. Used in combination, the single observation and annual triggers provide power over 90% to detect substantial increases above baseline at least once during five years (Table 3.1a), although in practice the power may be smaller for reasons discussed above under (6).

[1]As explained in (5), EPA believes that the error rates and power of these triggers will be acceptable in practice because best management practices (BMPs) reduce discharge loadings substantially.

Table 3.1a. **Statistical Triggers as Modified for Final Regulation: Percentage of Discharges Declared to Exceed Baseline Level** (at least once during 5 years of simulated monthly monitoring) [1]					
Annual Trigger	Single-Observation Trigger	Shift from Baseline to Remining Period [2]			
		-0.5 σ	0	1 σ	+2 σ
NA	Method 1	10 %	33 %	89 %	99 %
Method 1 (Multiplier = 1.96)	NA	3 %	11 %	59 %	94 %
Method 1 (Multiplier = 1.96)	Method 1	12 %	39 %	93 %	100 %
Method 1 (Multiplier = 1.96)	Method 2	7 %	29 %	91 %	100 %
NA	Method 2	5 %	22 %	86 %	100 %
Method 2 ($\alpha = 0.001$)	NA	2 %	11 %	65 %	97 %
Method 2 ($\alpha = 0.001$)	Method 2	7 %	28 %	91 %	100 %
Method 2 ($\alpha = 0.001$)	Method 1	12 %	38 %	93 %	100 %

[1] Assumes monthly serial correlation of 0.5 for log(x), with x distributed lognormally. Percentages were rounded to the nearest 1%.
[2] The shift was scaled in terms of standard deviation units (σ = standard deviation)

3.2 Statistical Procedures for Calculating Limits from Baseline Data

The procedures to be used for establishing effluent limitations for pre-existing discharges at coal remining operations, in accordance with the requirements set forth in 40 CFR part 434, Subpart G; Coal Remining are presented below. The requirements specify that pollutant loadings of total iron, total manganese, total suspended solids, and net acidity in pre-existing discharges shall not exceed baseline pollutant loadings. The two alternative procedures described (Method 1 and Method 2) are applied to determine site-specific, baseline pollutant loadings, and to determine whether discharge loadings during coal remining operations have exceeded baseline loading. For each procedure, both a monthly (single-observation) test and an annual test are applied. In order to sufficiently characterize pollutant loadings during baseline determination and during each annual monitoring period, the regulations require that at least one sample result be obtained per month for a period of 12 months.

The calculations described are applied to pollutant loadings. Each loading value is calculated as the product of a flow measurement and pollutant concentration taken on the same date at the same discharge sampling point, using standard units of flow and concentration (to be determined by the permitting authority). For example, flow may be measured in cubic feet per second, concentration in milligrams per liter, and the pollutant loading calculated in pounds per year.

In the event that a pollutant concentration in the data used to determine baseline is lower than the daily maximum limitation established in Subpart C for active mine wastewater, the statistical procedures should not establish a baseline more stringent than the BPT and BAT effluent standards established in Subpart C. Therefore, if the total iron concentration in a baseline sample is below 7.0 mg/L, or the total manganese concentration is below 4.0 mg/L, the baseline sample concentration should be replaced with 7.0 mg/L and 4.0 mg/L, respectively, for the purposes of some of the statistical calculations. The substituted values should be used for all methods described in this section with the exception of the calculation of the interquartile range (R) in Method 1 for the annual trigger, and in Method 2 for the single observation trigger. The interquartile range (R) is the difference between the quartiles M_{-1} and M_1; these values should be calculated using actual loadings (based on measured concentrations) when they are used to calculate R. This should be done in order to account for the full range of variability in the data.

3.2.1 Method 1

Method 1 is a modification of the methodology used by the Commonwealth of Pennsylvania. Computational details appear in Figure 3.2a. Pennsylvania's monthly and annual average checks can be described as follows:

Monthly (or single-observation maximum) check: A tolerance interval is estimated for the baseline loadings (for $n < 17$, the smallest and largest observations define the interval endpoints). The baseline upper bound (usually the maximum baseline loading) is the value of interest. Two consecutive exceedances of the upper bound trigger weekly monitoring. Four consecutive exceedances during weekly monitoring trigger a treatment requirement. Thus, six exceedances must occur consecutively before a treatment requirement is triggered.

Annual average check: A robust, asymptotic estimator[2] of a 95 percent confidence interval for the median is calculated for the baseline period and post-baseline periods; if the post-baseline interval exceeds the baseline interval, an exceedance is declared. This estimate is based upon McGill, Tukey, and Larsen (1978).

[2]Because loadings data for pre-existing discharges are highly asymmetric, and annual means and medians are likely to be somewhat asymmetrically distributed, EPA used an asymptotic approximation to develop the confidence intervals for the annual averages. However, the approximation results in confidence intervals that are symmetric rather than asymmetric. Thus, this approximation is expected to be accurate only for very large samples, because their means are approximately normally distributed (by the Central Limit Theorem). EPA has used this approximation for smaller samples because it provides reasonably good performance as demonstrated in simulations (see U.S.E.P.A., 2001c).

3.2.2 Method 2

Similarly to Method 1, Method 2 consists of two checks: an upper limit on single observations and an annual test of the mean or median. Computational details of Method 2 are provided in Figure 3.2b. The single-observation limit is a nonparametric estimate of the 99th percentile of loadings, developed using baseline data. The annual test of the average or median employs the nonparametric Wilcoxon-Mann-Whitney test.

3.2.3 Accelerated Monitoring

For Methods 1 and 2, triggered or accelerated monitoring is applied after two consecutive exceedances of the Single Observation Trigger L. If this occurs, weekly sampling must be commenced immediately. After four weekly samples are collected, the results should be compared to the Single Observation Trigger L. If three or fewer of the weekly observations exceed L, then monthly sampling can be resumed. However, if all four weekly observations exceed L, the baseline pollution loading has been exceeded.

Accelerated monitoring (if used as a condition or option for determining non-compliance) guards against a declaration of non-compliance on the basis of a transient exceedance, and provides a means to demonstrate continuing exceedances. It guards against the possibility of instituting expensive remedial measures when there was no continuing exceedance of baseline conditions.

Figure 3.2a: **Method 1:** The single-observation trigger is applied to each new measurement; the annual test is applied once a year, using all measurements for the past year

x_i = pollutant loading measurement (product of flow, concentration, and conversion factor)

n = number of x_i results in the baseline dataset

1. Single-observation trigger

Order all n baseline measurements such that $x_{(1)}$ is the lowest value, and $x_{(n)}$ is the highest.

If n<17, then:

The single-observation trigger will equal the maximum baseline value, $x_{(n)}$.

If n>16 then:

Calculate the sample median (M) of the baseline events:

If n is odd, then M equals $x_{(n/2+1/2)}$.

If n is even, then M equals $0.5*(x_{(n/2)}+ x_{(n/2+1)})$.

Calculate M_1 as the median between M and the maximum $x_{(n)}$.

Calculate M_2 as the median between M_1 and $x_{(n)}$.

Calculate M_3 as the median between M_2 and $x_{(n)}$.

Calculate M_4 as the median between M_3 and $x_{(n)}$.

The single-observation trigger L equals M_4.

If, during remining, two successive monthly observations exceed L, proceed immediately to weekly monitoring for four weeks (four weekly samples). If, during weekly monitoring, all four observations exceed L, declare exceedance of the baseline distribution.

2. Annual test

Calculate M and M_1 as described above.

Calculate M_{-1} as the median between the minimum $x_{(1)}$ and the sample median.

Calculate $R = (M_1 - M_{-1})$.

The subtle trigger (T) is calculated as:

$$T = M + \left[\frac{(1.815 * R)}{\sqrt{n}}\right]$$

Calculate M' and R' for a year's data during re-mining.

Calculate T' = M' - (1.815*R')/(n').

If T' > T , conclude that the median loading during re-mining has exceeded the median loading during the baseline period, and declare an exceedance.

Figure 3.2b: **Method 2: All three tests or limits are applied.** The single-observation trigger is applied to each new measurement; the annual test is applied once a year, using all measurements for the past year

1. Single-observation trigger

Calculate M and M_1 as described in Method 1 (Figure 3.2a).
Calculate M_{-1} as the median between the minimum $x_{(1)}$ and the sample median.
Calculate $R = (M_1 - M_{-1})$.
Calculate the Single Observation Trigger as $L = M_1 + (3 * R)$

If, during remining, two successive monthly observations exceeds L, proceed immediately to weekly monitoring for four weeks (four weekly samples). If, during weekly monitoring, all four observations exceed L, declare exceedance of the baseline distribution.

2. Annual comparison [1]

Compare baseline year loadings with current annual loadings using the Wilcoxon-Mann-Whitney test [2] for two independent samples. Use a one-tailed test with alpha 0.001.

---

[1] Hirsch, R.M., and J.R. Stedinger. 1987. Plotting Positions for Historical Floods and Their Precision. Water Resources Research. Vol. 23, No.4:715-727.
[2] See Conover, W.J., 1980, Practical Nonparametric Statistics, 2nd ed., and other textbooks.

References

Brady, K.B.C., M.W. Smith, and J. Schueck (editors), 1998. *Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania.* Pennsylvania Department of Environmental Protection, publ. no. 5600-BK-DEP2256, October 1998.

Hawkins, J.W. 1994. *A statistical evaluation of remining abandoned coal mines to reduce the effluent contaminant load.* International Journal of Surface Mining, Reclamation and Environment 8: 101-109.

Hornberger, R.J. et al., 1990. *Acid Mine Drainage from Active and Abandoned Coal Mines in Pennsylvania.* Chapter 32 of Water Resources in Pennsylvania: Availability, Quality, and Management, Pennsylvania Academy of Science, pp. 432-451.

Loftis, J.C. and R.C.Ward, 1980. *Sampling Frequency Selection for Regulatory Water Quality Monitoring.* Water Resources Bulletin, Vol.16, No. 3, pp.501-507.

McGill, R., J.W. Tukey, and W.A. Larsen, 1978. *Variations of Box Plots.* The American Scientist, Vol. 32, No. 1, pp. 12-16.

U.S.E.P.A., 1993. *Statistical Support Document for Proposed Effluent Limitations Guidelines and Standards for the Pulp, Paper, and Paperboard Point Source Category.* EPA publ. no. 821/R-93-023.

U.S.E.P.A., 1999. Office of Water. *Coal Remining Database*: 61 State Data Packages, March 1999. (docket number DCN 3054)

U.S.E.P.A., 2001a. *Statistical Analysis of Abandoned Mine Drainage in the Assessment of Pollution Load.* EPA-821-B-01-014.

U.S.E.P.A., 2001b. "Serial Correlation of Coal Mine Discharge Loadings," memorandum in the rulemaking record (docket number DCN 3050).

U.S.E.P.A., 2001c. "Evaluation of Statistical Triggers," memorandum in the rulemaking record (docket number DCN 3051).

U.S.E.P.A., 2001d. "Distribution & Variability of Coal Mine Discharge Loadings," memorandum in the rulemaking record (docket number DCN 3049).

Section 4.0 Baseline Sampling Duration and Frequency

4.1 Power and Sample Size

Power, in the context of this discussion, quantifies the probability that a particular statistical test and sample size will indicate that the mean or median loading has increased over the baseline, given that it truly has increased some specified amount (see Table 4.1a). The test is designed to guard against incorrectly concluding that the mean or median has increased by setting alpha at a low value. The probability is *less than* or equal to α that a statistical test and sample size will incorrectly indicate that the mean or median loading has increased over the baseline, given that it has *not* increased. If there has been a decrease in loadings, the risk of such an incorrect decision will be considerably less than alpha.

EPA evaluated the power of the statistical triggers by simulating a 60-month monitoring program for 5000 discharges, and recording the frequency with which the triggers indicated that the remining loadings exceeded baseline (see Section 3.1). The evaluations led to a choice of statistical procedures that achieve acceptable power and a reasonable balance between rates of false alarms and correct alarms.

The error rates of statistical decision procedures will depend upon the number of measurements ("sample size") used. If the false positive rate (alpha) is held constant, the power (the ability to detect an increase in pollutant load) will necessarily decrease as sample size decreases. EPA's evaluation assumed monthly sampling, using twelve samples taken over one year to characterize the baseline level, and using twelve samples taken over each year to monitor pollutant loads during remining. The performance of the evaluated statistical procedures was shown to be just adequate to meet the detailed objectives set out in Section 3.1 (see also U.S.E.P.A., 2001c) when based upon measurements taken once a month. Therefore, if these procedures are applied to measurements taken less frequently than once a month, or are applied to fewer than twelve measurements per year (for annual triggers), the ability to detect an increase in pollutant load will necessarily be lower than intended.

4.2 Necessary Duration and Frequency of Sampling

Without an adequate duration and frequency of sampling, the statistical procedures could establish baseline levels that are either too low or too high. Baseline sample collection requirements protect both the remining operator and the environment. If baseline characterization of pre-existing pollutant discharges is inadequate (for example, if it is based on too few samples), there is a chance that an operator could consistently face noncompliance by discharging pollutant loadings above an underestimated baseline. In addition, there is the chance that environmental improvement could be jeopardized by allowing for pollutant loading discharges at high levels that still fall below an overestimated baseline. EPA believes that 12 monthly samples are the minimum to derive a statistically sound estimate of baseline (U.S.E.P.A. 2001b).

EPA has determined that the smallest acceptable number and frequency of samples is 12 monthly samples, taken consecutively over the course of one year. Twelve samples may provide less than the required power if autocorrelation is very high, if sampling duration is less than a year, or if the sampling interval is shortened (e.g., to one week) while the number of samples is not increased above 12. Therefore, EPA has required a minimum of 12 monthly samples to establish baseline.

One of the criteria for sample size is the ability to detect a change of one standard deviation above baseline loadings with reasonably high power. Discharge flows, concentrations, and loadings vary remarkably among monthly or weekly samples over the course of 1-4 years (Brady et al., 1998; EPA, 2001b). Sample coefficients of variation (CV, the ratio of standard deviation to mean) for iron loadings range from 0.62 to 2.7 for 80% of discharges (U.S.E.P.A., 2001d).

Sample CVs for manganese loadings ranged from 0.54 to 1.7 for 80% of discharges (U.S.E.P.A., 2001d). The median CV is about 1.0, thus the standard deviation is as large as the mean. Assuming that the CV remains constant at 1.0 from baseline to post-baseline, an increase of one standard deviation above baseline means that the mean loading has doubled. Thus, it is important to have a sampling frequency and duration that will permit the statistical procedures to detect increases in loadings with high probability when the standard deviations increase.

A permitting authority may want to consider requiring more than twelve samples per year during and after the baseline year in order to increase power and in order to provide a fair chance of observing a representative sample of discharge flows and loadings.

It is possible that one year of sampling may not adequately characterize baseline pollutant levels, because discharge flows can vary among years in response to inter-year variations in rainfall and ground water flow. There is some risk that the particular year chosen to characterize baseline flows and loadings will be a year of atypically high or low flow or loadings. Permitting authorities should be aware of this risk and may want to inform permittees of this risk in order to encourage multi-year characterization of baseline. To design a procedure to evaluate inter-year variations, EPA evaluated correlations between discharge flow and various parameters of existing mine discharge data and indices for which data spanning over many years are available to the public (i.e., Palmer Indices, Standardized Precipitation Index, Crop Moisture Index, Surface Water Supply Index, and USGS Current and Historical Daily Streamflow). EPA concluded that historical stream flow data from a USGS gage station associated with a discharge could be used to test whether the given baseline year was significantly different from the previous years. This would be done by comparing the mean stream flow for the baseline year to the 2.5^{th} and 97.5^{th} percentiles of annual mean stream flows prior to the baseline year. If the mean stream flow for the baseline year falls below the 2.5^{th} percentile or above the 97.5^{th} percentile, the year may have unusually low or high flow, respectively. In such cases, it may be best to continue baseline sampling for another year.

A sampling plan should be designed to prevent biased sampling. Sampling, during and after baseline, should systematically cover all periods of the year during which substantially high or low discharge flows can be expected. Unequal sampling of months could bias the baseline mean or median toward high or low loadings by over sampling of high-flow or low-flow months. However, unequal sampling of different time periods can be accounted for by using statistical estimation procedures appropriate to stratified sampling. Stratified seasonal sampling, possibly with unequal sampling of different time periods, is a suitable alternative to regular monthly sampling, provided that correct statistical estimation procedures for stratified sampling are applied to estimate the mean, median, variance, interquartile range, and other quantities used in the statistical procedures, and provided that at least one sample is taken per month over the course of one year.

Flow measurement methods also should accurately measure flows during high-flow events. If the discharge overflows or bypasses the weir or flume, or if a measurement is not made as scheduled on a high-flow day, statistical characterizations of flow and loading will be inaccurate. The sampling location and methods should be designed as much as possible to permit access and sampling on all scheduled days, and to avoid the need to reschedule sampling because flow is extremely high.

References

Brady, K.B.C., R.J. Hornberger, and G. Fleeger, 1998. Influence of Geology on Postmining Water Quality: Northern Appalachian Basein. Chapter 8 in Coal Mine Drainage Prediction abnd Pollution Prevention in Pennsylvania. Edited by K.B.C. Brady, M.W. Smith, and J. Schueck, Department of Environmental Protection, pp. 8-1 to 8-92.

Griffiths, J.C., 1990. Letter to M. Smith, Pennsylvania Department of Environmental Resources dated January 14, 1990, 3 pages, with attachment of 54 pages titled "Chapter 3, Statistical Analysis of Mine Drainage Data".

Hawkins, J.W., 1994. Statistical Characteristics of Coal Mine Discharges on Western Pennsylvania Remining Sites. Water Resources Bulletin, Vol. 30, No. 5, pp. 861-869.

McGill, R., J.W. Tukey, and W.A. Larsen, 1978. Variations of Box Plots. The American Scientist, Vol. 32, No. 1, pp. 12-16.

Millard, S.P., 1998. Environmental Stats for S-Plus: Users Manual for Windows and Unix. Springer-Verlag: New York, NY.

Sanders T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian, and V. Yevjevich, 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, CO.

U.S. EPA, 2001a. Coal Remining Best Management Practices Guidance Manual. EPA-821-B-01-010.

U.S.E.P.A., 2001b. *Statistical Analysis of Abandoned Mine Drainage in the Assessment of Pollution Load.* EPA-821-B-01-014.

U.S.E.P.A., 2001c. "Evaluation of Statistical Triggers," memorandum in the rulemaking record (docket number DCN 3051).

U.S.E.P.A., 2001d. "Distribution & Variability of Coal Mine Discharge Loadings," memorandum in the rulemaking record (docket number DCN 3049).

Section 5.0 Long-term Monitoring Data and Case Studies

A remining study conducted by EPA and PA DEP from 1984 through 1988 involved the statistical analysis of long-term abandoned mine discharge data from six sites in Pennsylvania. This study is described in Section 1.0 of this document. These sites and corresponding discharge data were selected because they contained a sufficient number of samples for examining mine drainage discharge behavior with univariate, bivariate, and time series statistical analyses following the algorithm shown in Figure 1.2a. The results of the statistical analyses are included in a series of eight unpublished reports prepared for EPA and PA DEP by Dr. J. C. Griffiths of the Pennsylvania State University. These reports are discussed in EPA's Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits (USEPA, 2001; EPA-821-B-01-014).

Sections 3.0 and 4.0 of this document describe the statistical methodology for establishing baseline for pre-existing discharges, and determined that the minimum baseline sampling duration and frequency is twelve samples in one year at approximately monthly intervals. Some discharge datasets in Pennsylvania contain more than twelve samples. These additional samples represent pre-mining baseline conditions of more than one year, and in some cases, discharges were monitored more frequently than monthly (e.g., weekly). In this Section, data from seven discharges at six sites previously studied by Griffiths are further examined by varying the baseline sampling interval and the number of samples used to establish baseline.

A benefit of further evaluation of the EPA/PA DEP remining study is that for some of the six sites, there are now approximately ten years of additional monitoring data. In addition, PA DEP has issued approximately three hundred remining permits since 1985, and for many completed sites there are complete historical records of discharge data from pre-mining baseline conditions, through active surface mining phases (open pit), to post-mining reclamation. Several examples of these long-term monitoring case studies are presented in this Section. These studies provide additional information on the magnitude and variability of natural seasonal variations, and show mining-induced changes in water quality and pollution load.

The long-term monitoring case studies in this Section can be used as examples of the application of the baseline statistical test described in Section 3.0 to actual remining datasets. The quality control approach to long-term baseline monitoring data is presented in the figures included in Section 5.3. Further examples of application of the baseline statistical test are presented in Appendix A.

5.1 A Comparison of Seven Long-term Water Quality Datasets

As part of the investigation documenting the baseline pollution load of pre-existing acid mine drainage (AMD) discharges, seven individual discharges with long-term water quality records were studied. Each of these seven discharges had datasets of at least 3 years duration, and were sampled at least monthly and as frequently as weekly. The seven discharges represent the three principal discharge behavior types (typical, slug, steady) discussed in Section 2.0. Table 5.1a lists the discharge behavior type, location, period of record, and number of samples for each of the seven long-term discharges evaluated.

Table 5.1a: Long Term Acid Mine Drainage Datasets

Dataset	Discharge Behavior Type	Location	Period of Record	Number of Samples
Arnot-3	typical	Tioga County, PA	1980 - 1983	82
Arnot-4	typical	Tioga County, PA	1980 - 1983	81
Clarion	typical	Clarion County, PA	1982 - 1986	79
Ernest	slug	Indiana County, PA	1981 - 1984	189
Fisher	typical	Lycoming County, PA	1982 - 1985	36
Hamilton	typical	Centre County, PA	1981 - 1985	109
Markson	steady	Schuylkill County, PA	1984 - 1986	99

The discharge behaviors discussed in Section 2.0 are summarized as:

1) Typical Discharge Response: Typical discharge response exhibits lower pollutant concentrations during high-flow periods and higher concentrations during low-flow periods. Most pre-existing discharges exhibit this type of behavior. These discharges tend to vary significantly, both seasonally and in response to individual recharge events. Five of the seven discharges listed in Table 5.1a exhibit this characteristic flow-response behavior (Arnot-3, Arnot-4, Clarion, Hamilton, and Fisher). All but the Clarion discharge are from relatively small (less than one square mile) underground mine complexes. The Clarion discharge emanated from a previously surface mined area.

2) Slug Response: The Ernest discharge emanates from an extensive unreclaimed coal refuse pile and exhibits highly variable behavior responding to individual precipitation events. It exhibits "slugger response" behavior. Increases in flow are not necessarily offset by decreased concentration and at times may even exhibit increased concentration due to the build-up of water-soluble acid salts in the unsaturated zone during periods of decreased precipitation or little recharge. These discharge types are extremely variable in flow and pollutant loading rates. They present the biggest challenge for accurate documentation of baseline pollution load.

3) Steady response: The Markson discharge illustrates steady response behavior typical of discharges from very large underground mine pools. These discharges vary seasonally, but because of their large ground water storage capacity, respond in a damped fashion and do not exhibit large changes in pollutant concentrations. These types of discharges are the least variable in terms of baseline pollution load. However, because loading rates change slowly, they are also the most susceptible to year-to-year variation in pollution load.

Baseline pollution load statistical summaries were calculated for each dataset using the exploratory data analysis approach discussed in Section 3.0. It is rare, however, that datasets of this duration and with as great a number of samples are available. Coal remining operations tend to be relatively small and run by small mining companies. The time available for a small coal operator to lease a reminable reserve, gather permit application information, obtain a permit, and actually mine and reclaim a site is frequently very short, making long-term baseline sampling periods infeasible. Moreover, because these operations tend to be economically marginal, large sample sizes with frequent sampling intervals can be cost-prohibitive. In view of these constraints, the primary concern with establishing a valid baseline is to determine the minimum sampling period and sampling interval which will yield statistically valid results.

The problem of determining the minimum number of samples and minimum sampling period that would yield statistically valid results was examined using the long-term datasets listed in Table 5.1a. It was first assumed that the baseline pollution load determined using all of the available samples over the entire period of record represents the most accurate baseline achievable. Reduced datasets (subsets) were then used to recalculate the baseline, and comparisons to the full dataset were made using the following data subsets: monthly sample collection, quarterly sample collection, and nine-month sample collection (February through October, excluding November, December, and January). The nine-month sample collection subset was used to test the possibility that excluding three months (typically November, December, and January are average flow months) could adequately represent the full water year. This comparison is presented in Table 5.1b. In addition, baselines were calculated for each full calendar year (full data baselines) to examine the extent of year-to-year variability in baseline pollution load (Table 5.1c).

For simplicity, this evaluation looks at net acidity (the principal parameter of concern and indicator of pH) and iron (the most prevalent metal present in AMD). Tables 5.1b and 5.1c list median loads and calculated approximate 95 percent confidence intervals (C.I.) around each median load. Assuming that the full data baseline best represents the true population median load, the percent error for each data subset is calculated as the difference between the full data baseline value and the subset baseline value, divided by the full data baseline value. Percent errors are presented in Table 5.1b. While no particular percent error is considered to be acceptable or unacceptable, these percentages are useful for examining which subsets provide the closest approximations of the full data baseline. Percent errors less than 10 are highlighted. Table 5.1c presents the median baseline pollution loads and 95 percent confidence intervals for each of the seven discharges studied. Years that do not show overlapping 95 percent confidence intervals are considered to be statistically different at the 95 percent significance level.

Table 5.1b: Comparison of Median Acidity and Iron Loads by Sample Period and Interval

Parameter	Full Data	9 Month Data	Percent Error	Monthly Samples	Percent Error	Quarterly Samples	Percent Error
Arnot-3							
Number of Samples	82	66		43		14	
Median Acid Load	72.3	84.1	16.32 %	73.9	2.21 %	72.3	0.00 %
Upper 95% C.I.	86.53	101.59		91.09		102.71	
Lower 95% C.I.	58.07	66.61		56.71		41.89	
Median Iron Load	0.96	1.17	21.88 %	0.95	-1.04 %	0.96	0.00 %
Upper 95% C.I.	1.26	1.55		1.27		1.46	
Lower 95% C.I.	0.66	0.79		0.63		0.46	
Arnot-4							
Number of Samples	81	66		43		14	
Median Acid Load	194	221	13.92 %	193	-0.52 %	185	-4.64 %
Upper 95% C.I.	232.31	263.22		233.41		248.05	
Lower 95% C.I.	155.69	178.78		152.59		121.95	
Median Iron Load	2.70	3.00	11.11 %	2.50	-7.41 %	2.60	-3.70 %
Upper 95% C.I.	3.35	3.78		3.22		3.62	
Lower 95% C.I.	2.05	2.22		1.78		1.58	
Clarion							
Number of Samples	75	53		41		28	
Median Acid Load	39.50	40.00	1.27 %	39.00	-1.27 %	40.00	1.27 %
Upper 95% C.I.	49.11	51.45		52.01		54.74	
Lower 95% C.I.	29.89	28.55		25.99		25.26	
Median Iron Load	5.51	4.26	-22.69 %	4.45	-19.24 %	7.37	33.76 %
Upper 95% C.I.	7.27	6.07		7.02		10.51	
Lower 95% C.I.	3.75	2.45		1.88		4.23	
Ernest							
Number of Samples	189	146		53		19	
Median Acid Load	1456	1682	15.52 %	2048	40.66 %	1882	29.26 %
Upper 95% C.I.	1991.91	2410.88		2923.35		3805.81	
Lower 95% C.I.	920.09	953.12		1172.65		-41.81	
Median Iron Load	229	264	15.28 %	304	32.75 %	348	51.97 %
Upper 95% C.I.	342.83	412.68		474.61		662.48	
Lower 95% C.I.	115.17	115.32		133.39		33.52	
Fisher							

Parameter	Full Data	9 Month Data	Percent Error	Monthly Samples	Percent Error	Quarterly Samples	Percent Error
Number of Samples	35	24		24		10	
Median Acid Load	72	82	13.89 %	85	18.06 %	102	41.67 %
Upper 95% C.I.	95.00	109.47		121.73		165.38	
Lower 95% C.I.	49.00	54.53		48.27		38.62	
Median Iron Load	1.4	1.4	0.00 %	1.4	0.00 %	1.4	0.00 %
Upper 95% C.I.	1.74	1.75		1.66		1.63	
Lower 95% C.I.	1.06	1.05		1.14		1.17	
Hamilton-8							
Number of Samples	109	85		52		38	
Median Acid Load	59.00	66.86	13.32 %	58.70	-0.51%	55.70	-5.59 %
Upper 95% C.I.	67.92	77.16		68.90		72.60	
Lower 95% C.I.	50.80	56.56		48.50		38.30	
Median Iron Load	2.66	3.12	17.29 %	2.63	-1.13 %	1.81	-31.95 %
Upper 95% C.I.	3.19	3.76		3.45		2.77	
Lower 95% C.I.	2.13	2.48		1.81		0.85	
Markson							
Number of Samples	98	77		30		22	
Median Acid Load	1467	1452	-1.02 %	1491	1.64 %	1546	5.39 %
Upper 95% C.I.	1575.47	1597.55		1624.02		1816.11	
Lower 95% C.I.	1358.53	1306.45		1357.98		1275.89	
Median Iron Load	408	402	-1.47 %	402	-1.47 %	402	-1.47 %
Upper 95% C.I.	430.76	428.18		428.14		434.13	
Lower 95% C.I.	385.24	375.82		375.56		369.87	
Average of All Discharges							
Median Acid Load			10.75 %		9.27 %		12.55 %
Median Iron Load			12.82 %		9.01 %		17.55 %

Table 5.1c: Comparison of Median Acidity and Iron Loads by Baseline Sampling Year

Parameter	Full Data	1980	1981	1982	1983	1984	1985
Arnot-3							
Number of Samples	82	17	21	27	15		
Median Acid Load	72.3	66.9	63.8	83.9	86.5		
Upper 95% C.I.	86.53	94.23	76.79	110.69	128.53		
Lower 95% C.I.	58.07	34.37	46.62	43.72	39.51		
Median Iron Load	0.96	0.57	0.98	1.44	1.17		
Upper 95% C.I.	1.26	0.90	1.37	2.04	2.12		
Lower 95% C.I.	0.66	0.24	0.59	0.84	0.22		
Arnot-4							
Number of Samples	81	17	20	29	15		
Median Acid Load	194	157	159	208	242		
Upper 95% C.I.	232.31	253.83	209.32	256.19	368.52		
Lower 95% C.I.	155.69	60.17	108.68	159.81	115.48		
Median Iron Load	2.7	1.5	1.6	3.0	3.0		
Upper 95% C.I.	3.35	2.95	1.76	4.11	5.20		
Lower 95% C.I.	2.05	0.05	1.44	1.89	0.80		
Clarion							
Number of Samples	75	17	20	11	16	9	
Median Acid Load	39.5	41.0	56.5	27.0	14.0	42.0	
Upper 95% C.I.	49.11	74.00	75.80	46.41	27.99	61.26	
Lower 95% C.I.	29.89	8.00	37.20	7.59	0.01	22.74	
Median Iron Load	5.51	4.13	10.78	7.65	1.66	5.69	
Upper 95% C.I.	7.27	7.19	14.49	18.00	3.13	9.68	
Lower 95% C.I.	3.75	1.07	7.07	-2.70	0.19	1.70	
Ernest							
Number of Samples	189	16	38	47	49	39	
Median Acid Load	1456	1736	615	574	5193	1697	
Upper 95% C.I.	1991.91	2742.88	1141.38	1295.70	6551.26	2906.26	
Lower 95% C.I.	920.09	729.12	88.62	-147.70	3834.74	487.74	
Median Iron Load	229	225	85	60	1069	216	
Upper 95% C.I.	342.83	327.04	169.49	142.37	1346.84	448.62	
Lower 95% C.I.	115.17	122.96	0.51	-22.37	791.16	-16.62	
Fisher							
Number of Samples	35	9	8	17	21	12	8
Median Acid Load	72	49	101	80	36	26	42
Upper 95% C.I.	95.00	86.76	202.90	119.39	45.26	41.20	60.10
Lower 95% C.I.	49.00	11.24	-0.90	40.61	26.74	10.80	23.90
Median Iron Load	1.4	1.5	2.1	1.2	0.9	0.2	0.2
Upper 95% C.I.	1.74	2.13	3.65	1.59	1.12	0.41	0.36
Lower 95% C.I.	1.06	0.87	0.55	0.81	0.68	-0.01	0.04
Hamilton-8							
Number of Samples	109	16	24	27	25	17	
Median Acid Load	59.00	56.70	69.10	37.60	54.54	77.40	

Parameter	Full Data	1980	1981	1982	1983	1984	1985
Upper 95% C.I.	67.92	79.01	87.02	60.69	71.41	91.27	
Lower 95% C.I.	50.80	34.39	51.18	14.51	37.67	63.53	
Median Iron Load	2.66	4.35	3.50	1.07	1.53	3.34	
Upper 95% C.I.	3.19	6.74	4.98	1.72	2.16	4.21	
Lower 95% C.I.	2.13	1.96	2.02	0.42	0.90	2.47	
Markson							
Number of Samples	98	15	49	34			
Median Acid Load	1467	1502	1327	1888			
Upper 95% C.I.	1575.47	1726.37	1445.08	2366.21			
Lower 95% C.I.	1358.53	1277.63	1208.92	1409.79			
Median Iron Load	408	336	403	449			
Upper 95% C.I.	430.76	406.26	423.90	512.73			
Lower 95% C.I.	385.24	265.74	382.10	385.27			

5.1.1 Sampling Interval

For net acidity loads, two (Ernest and Fisher) out of the seven discharge datasets exceeded 10 percent error when both monthly and quarterly sample intervals were used. The average error for net acidity load using monthly sample collection was 9.27 percent. The average error for net acidity load using quarterly samples was 12.55 percent. For iron loads, 10 percent error was exceeded for monthly sampling on the Clarion and Ernest discharges. Ten percent error was exceeded with quarterly sampling for the Clarion, Ernest, and Hamilton discharges. The average error for iron load was 9.01 percent for monthly samples and 17.55 percent for quarterly samples. Monthly sampling yielded results closer to the full baseline than quarterly sampling. The effect of quarterly sampling would likely be even more pronounced if a shorter sampling period (e.g., one year) had been used.

The difference in baselines calculated for each discharge using monthly samples versus quarterly samples is illustrated in Figures 5.1a through 5.1ab. These figures also present yearly comparison of baseline pollutant loadings. In the data comparison (monthly versus quarterly) figures, the short horizontal lines represent the median values. The parallel vertical lines represent the range of the 95 percent confidence intervals around the median. The left-hand line shows the 95 percent confidence interval calculated based on the actual number of samples (N) as listed in Table 5.1b. However, because each sample subset contains a different number of samples, the confidence intervals are affected by different N values. A smaller number of samples results in a wider confidence interval. For purposes of comparison between datasets, the right-hand line shows the 95 percent confidence interval based on an arbitrarily set value for N equal to 12.

Figure 5.1a: Arnot-3 Acidity Loading (1980-1981)

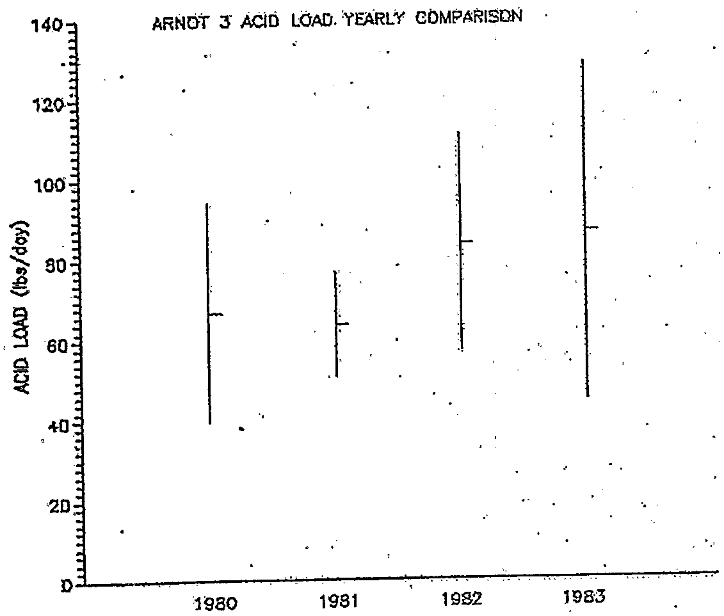


Figure 5.1b: Arnot-3 Flow Data Comparison

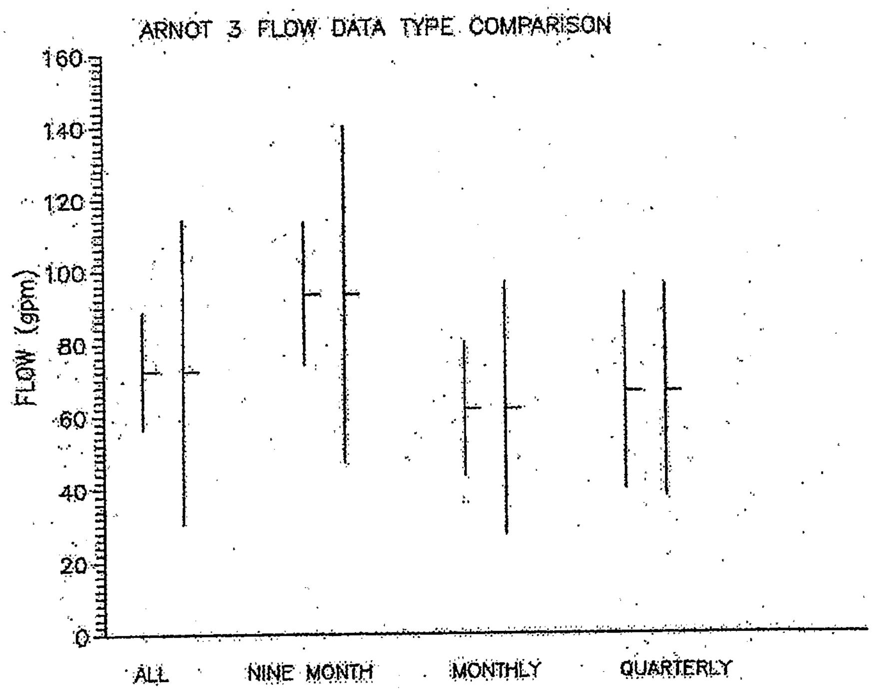


Figure 5.1c: Arnot-3 Iron Load Data Comparison

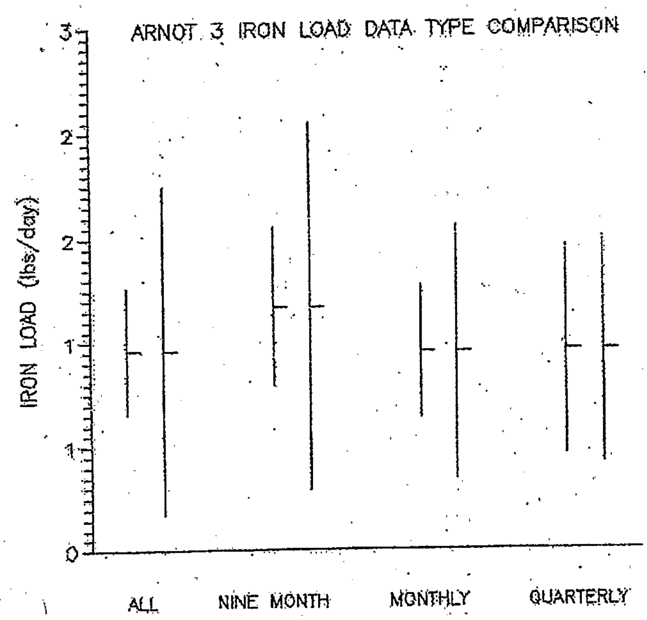


Figure 5.1d: Arnot-3 Acidity Load Data Comparison

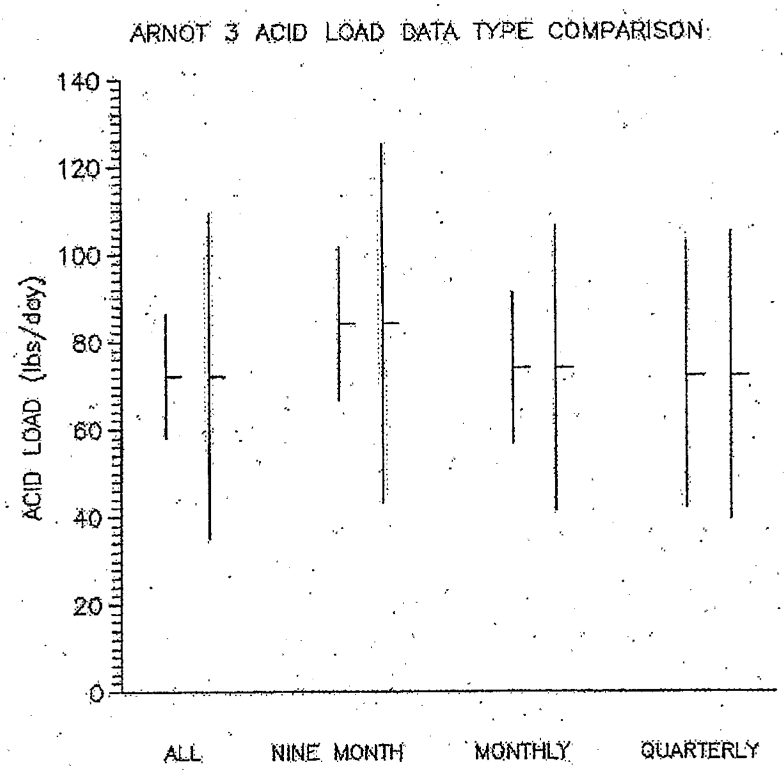


Figure 5.1e: Arnot-3 Monthly Flow Comparison

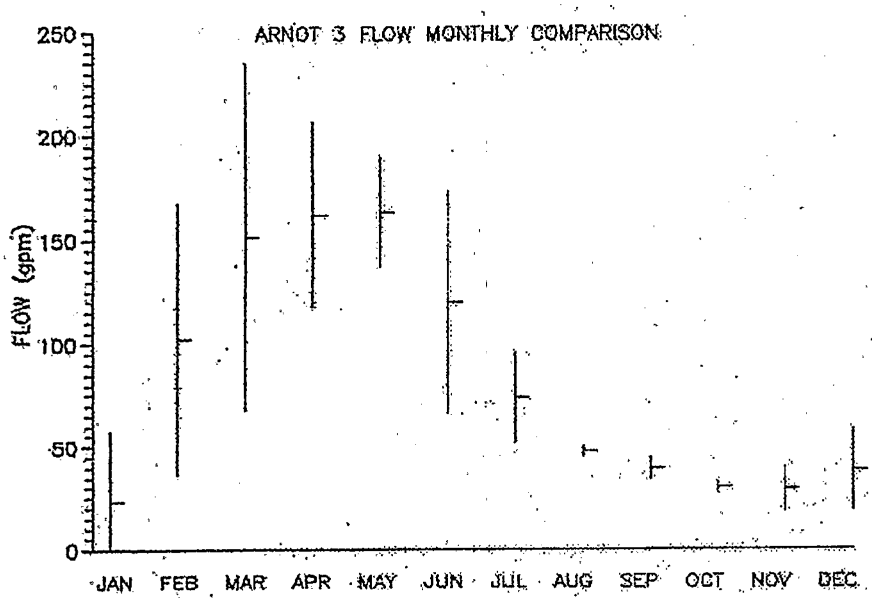


Figure 5.1f: Arnot-3 Monthly Acidity Load Comparison

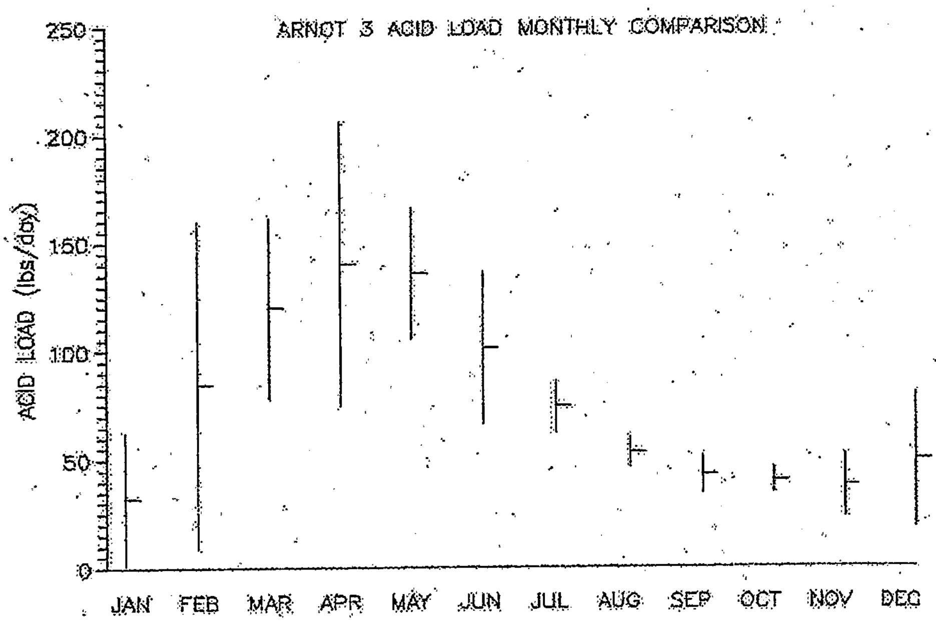


Figure 5.1g: Arnot-4 Acidity Load Data Comparison

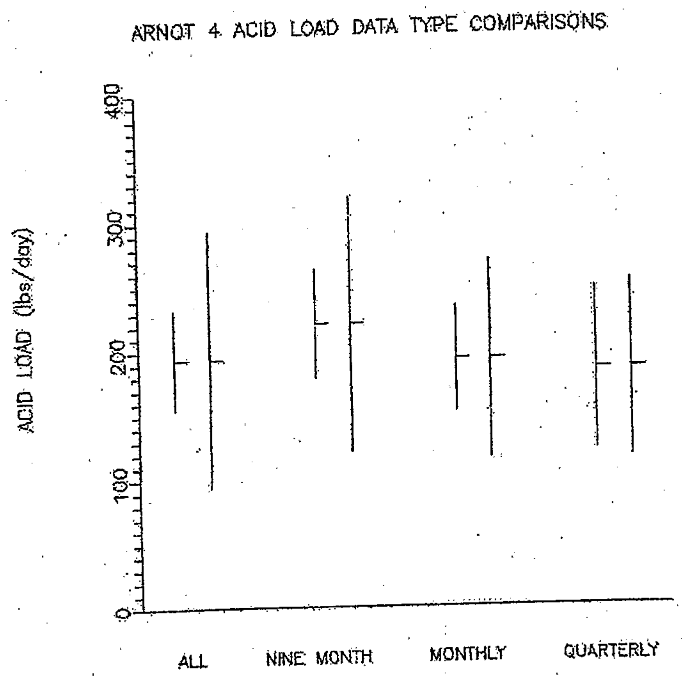


Figure 5.1h: Arnot-4 Acidity Load (1980-1983)

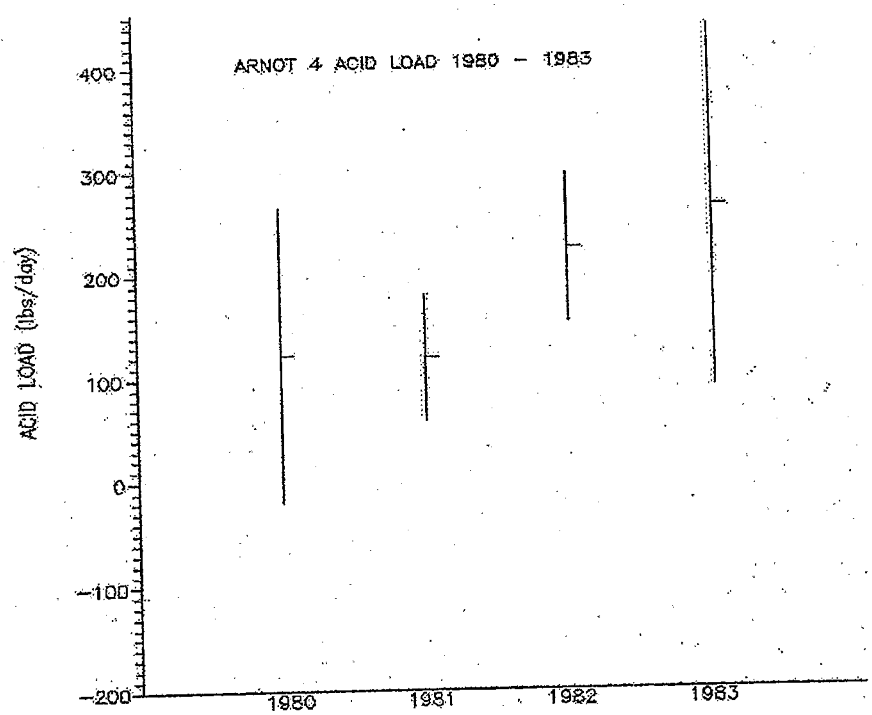


Figure 5.1i: Clarion Acidity Load Data Comparison

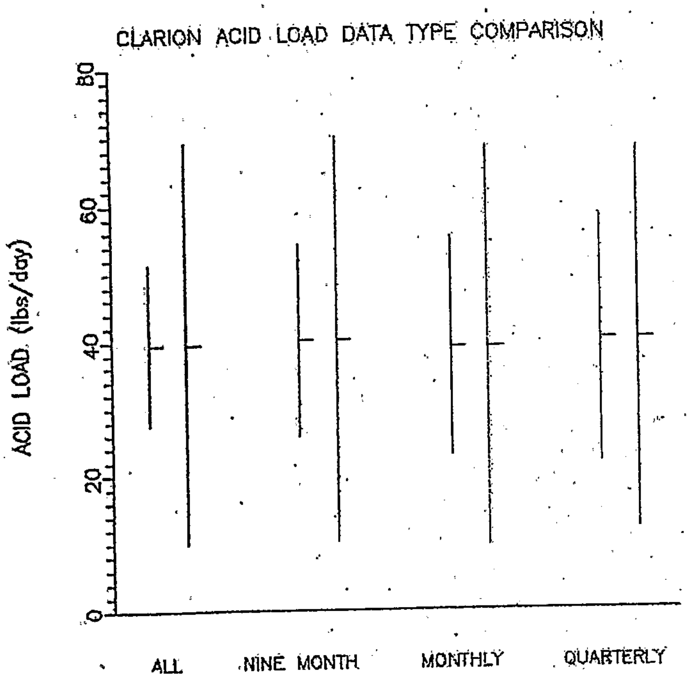


Figure 5.1j: Clarion Iron Load Data Comparison

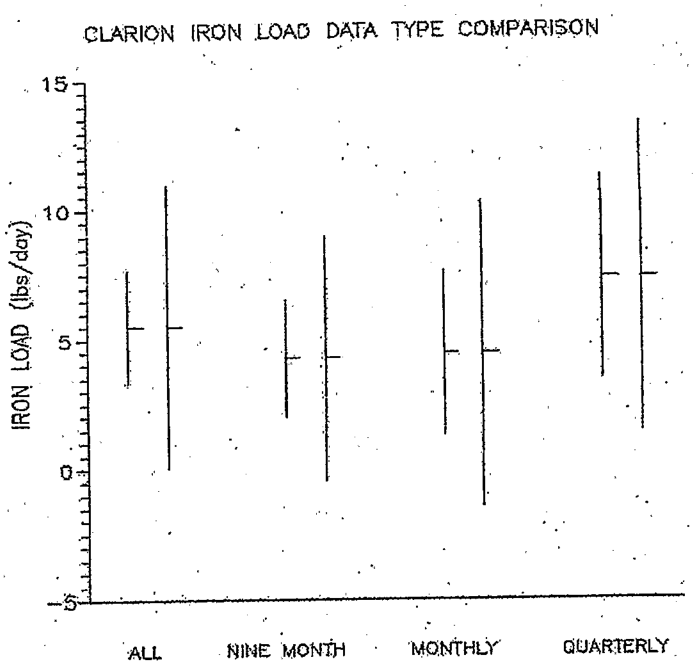


Figure 5.1k: Clarion Acidity Load (1982-1986)

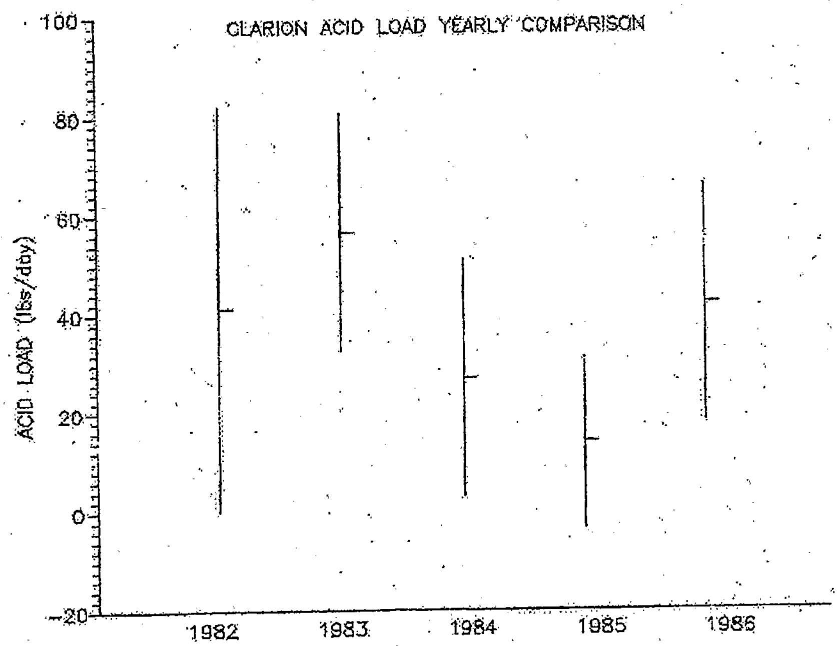


Figure 5.1l: Clarion Iron Load (1980-1983)

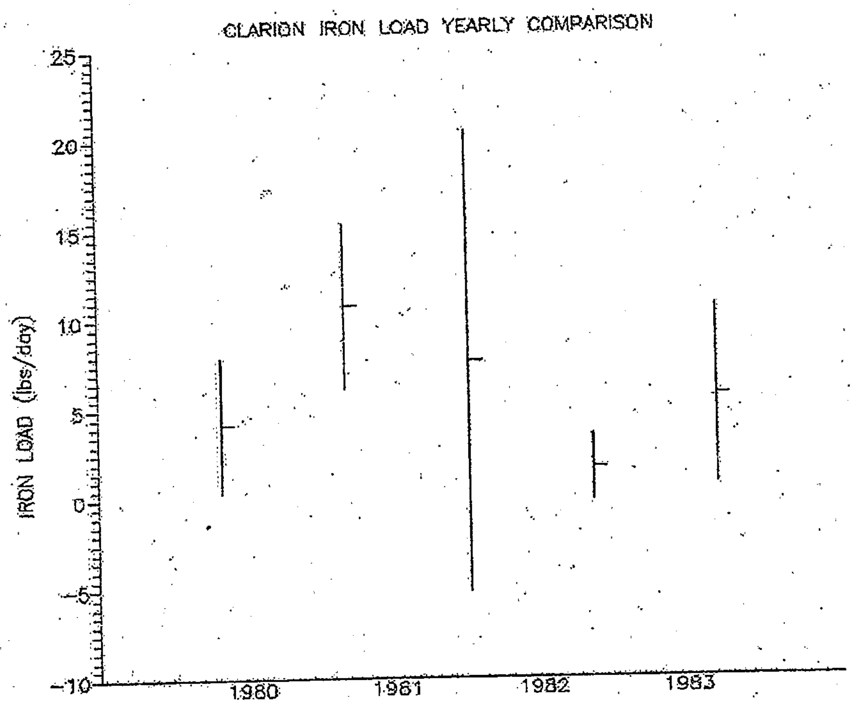


Figure 5.1m: Ernest Acidity Load Data Comparison

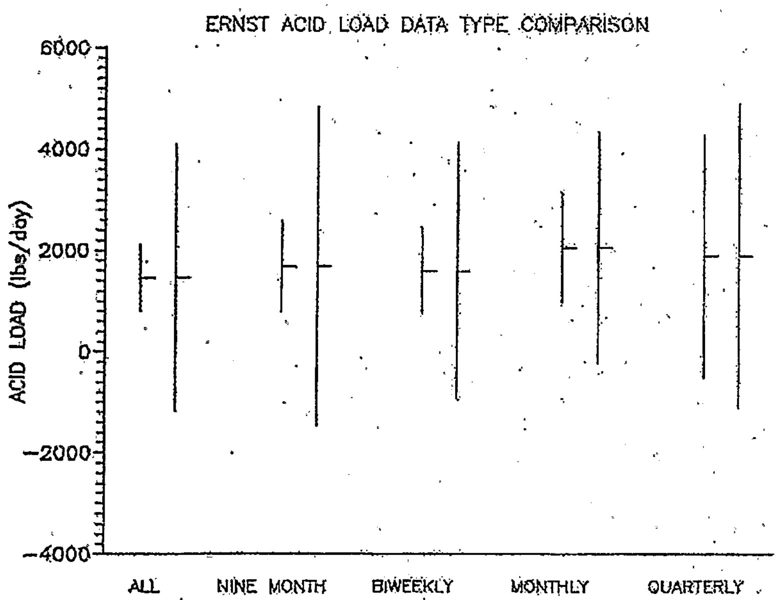


Figure 5.1n: Ernest Iron Load Data Comparison

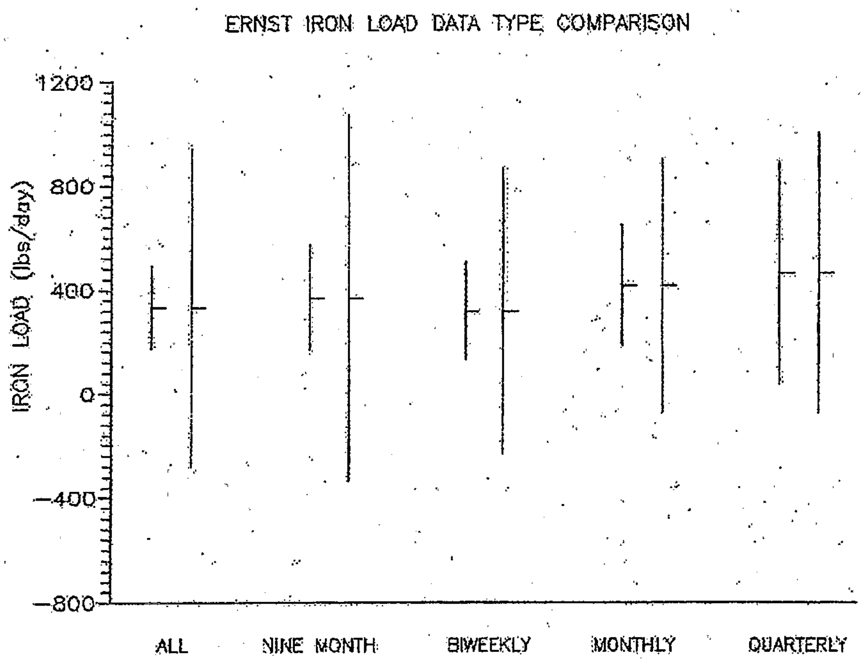


Figure 5.1o: Ernest Acidity Load Data Comparison

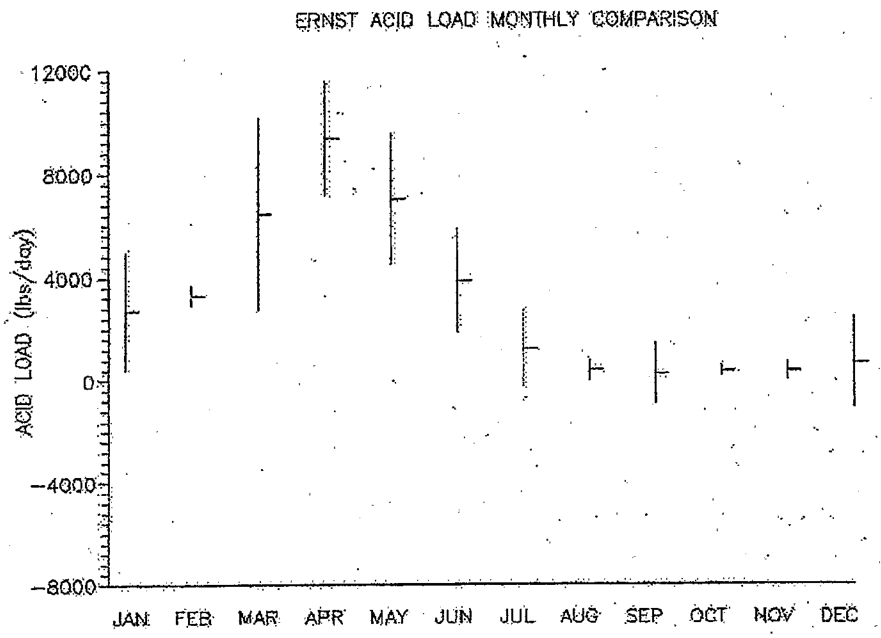


Figure 5.1p: Ernest Acidity Load (1981-1985)

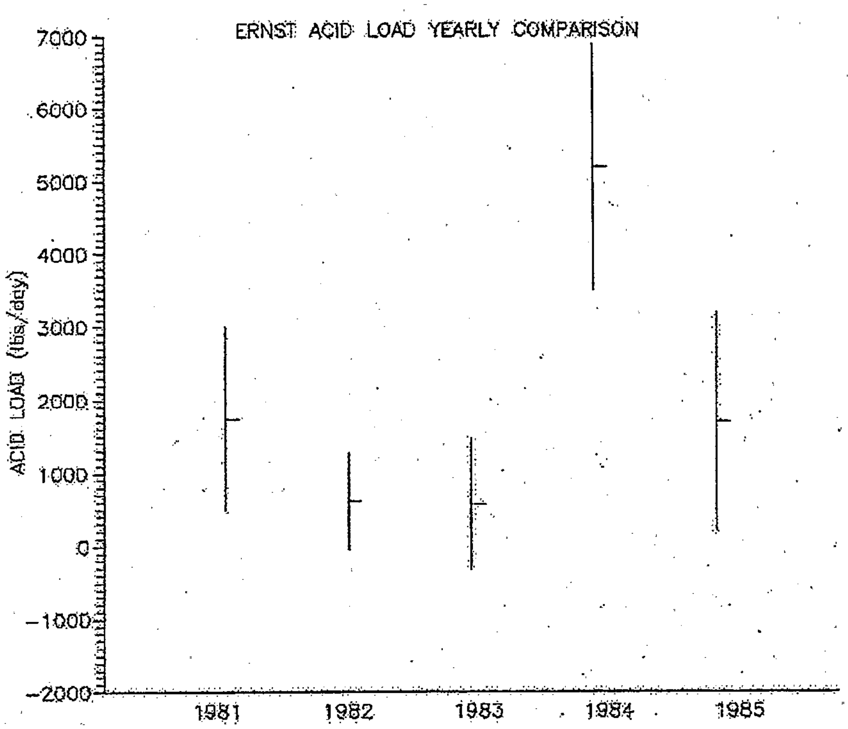


Figure 5.1q: Fisher Monthly Acidity Load

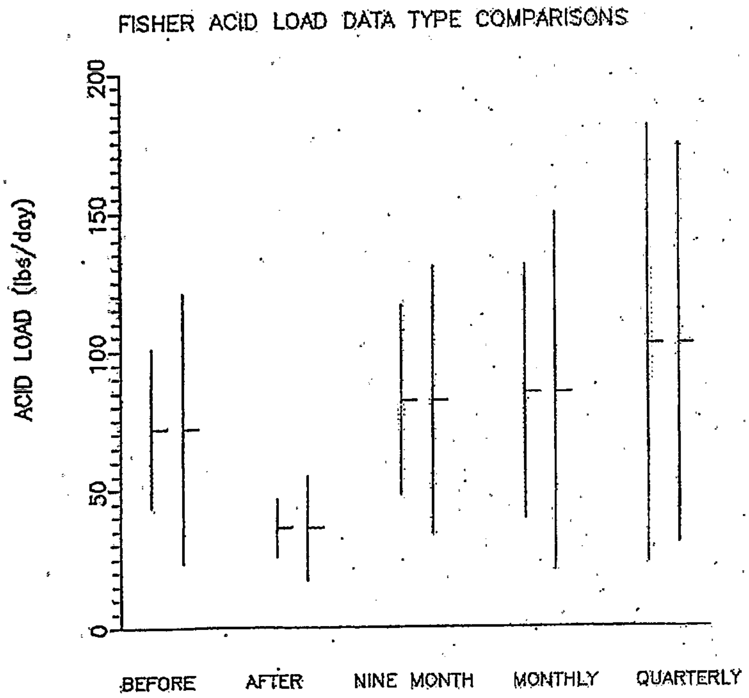


Figure 5.1r: Fisher Iron Load Data Comparison

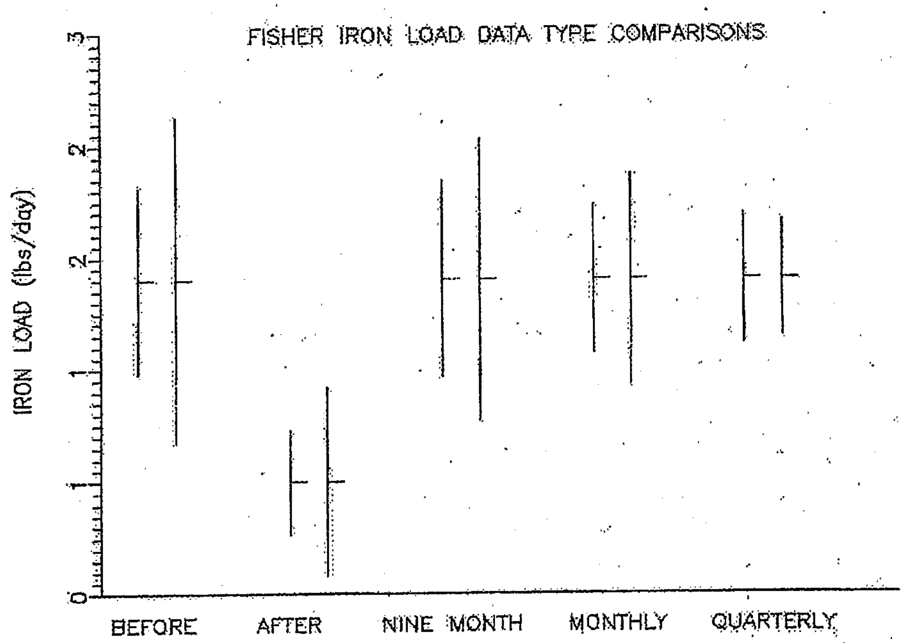


Figure 5.1s: Fisher Acidity Load Data (1982-1987)

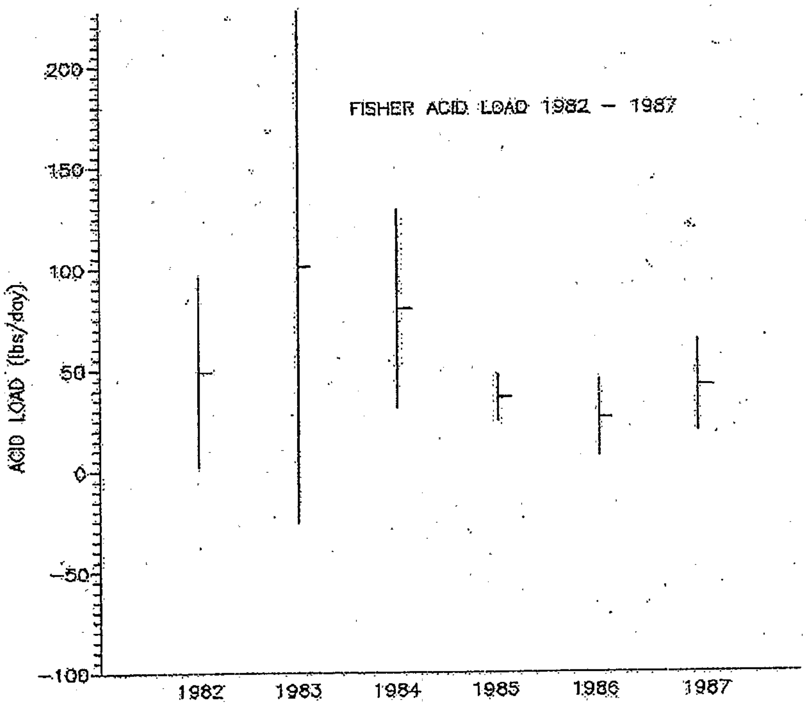


Figure 5.1t: Fisher Iron Load Data (1982-1987)

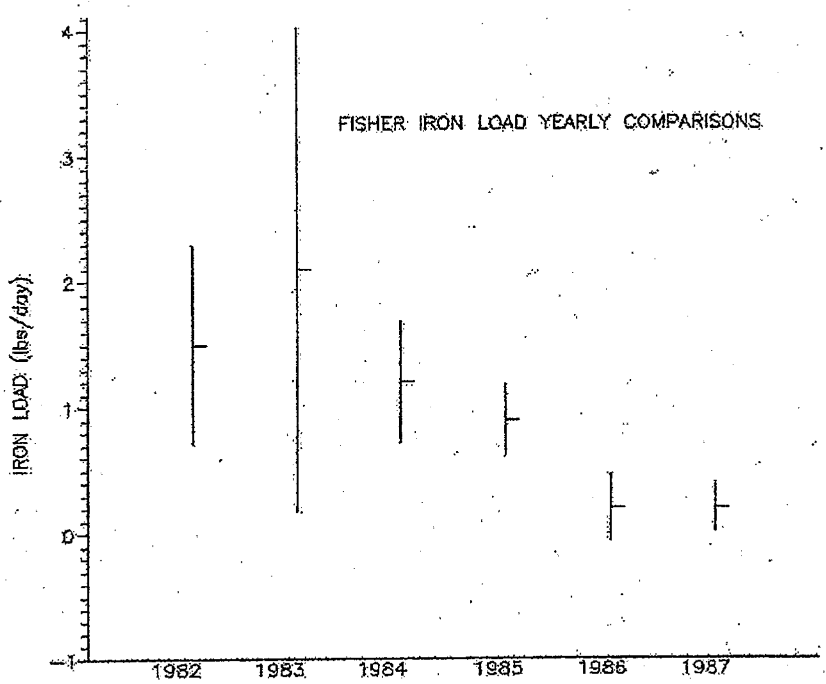


Figure 5.1u: Hamilton-8 Acidity Load Data Comparison

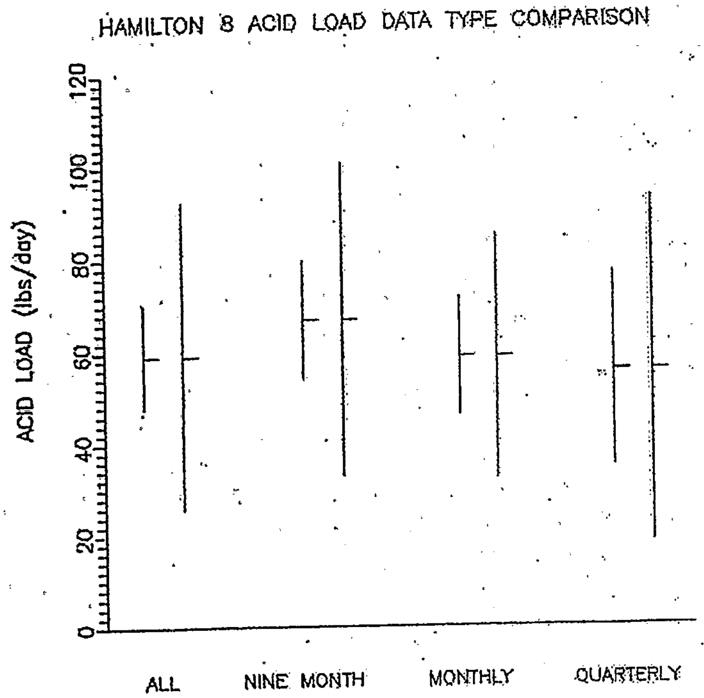


Figure 5.1v: Hamilton-8 Iron Load Data Comparison

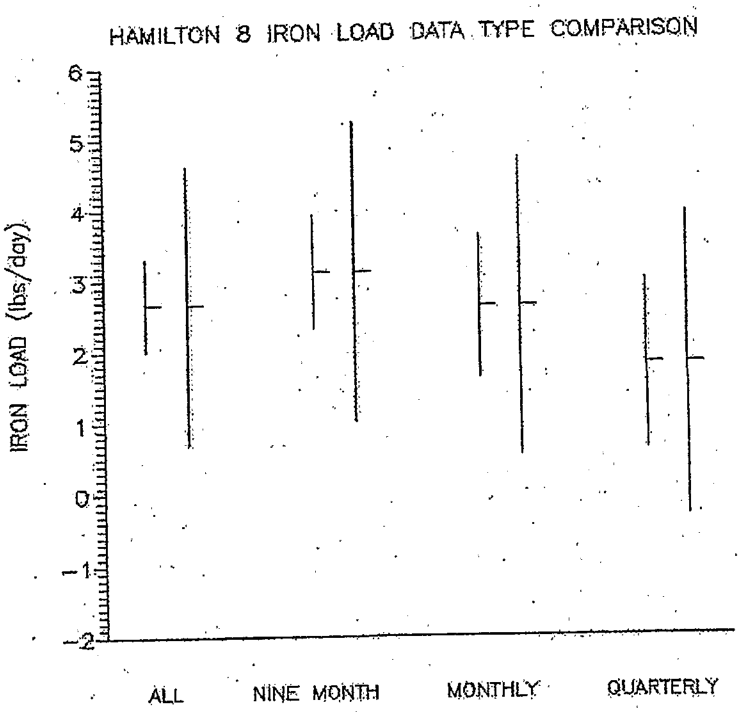


Figure 5.1w: Hamilton-8 Acidity Load (1981-1985)

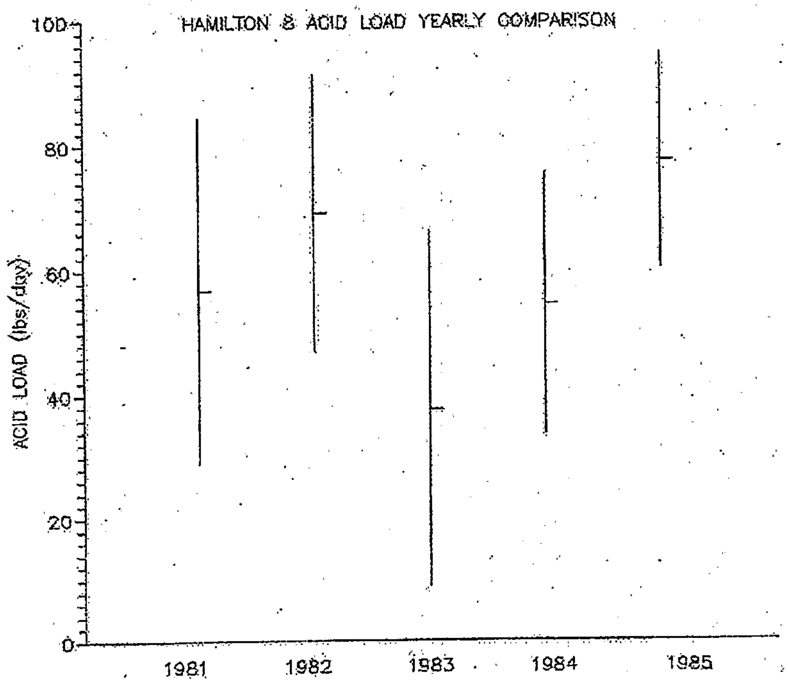


Figure 5.1x: Hamilton-8 Iron Load (1981-1985)

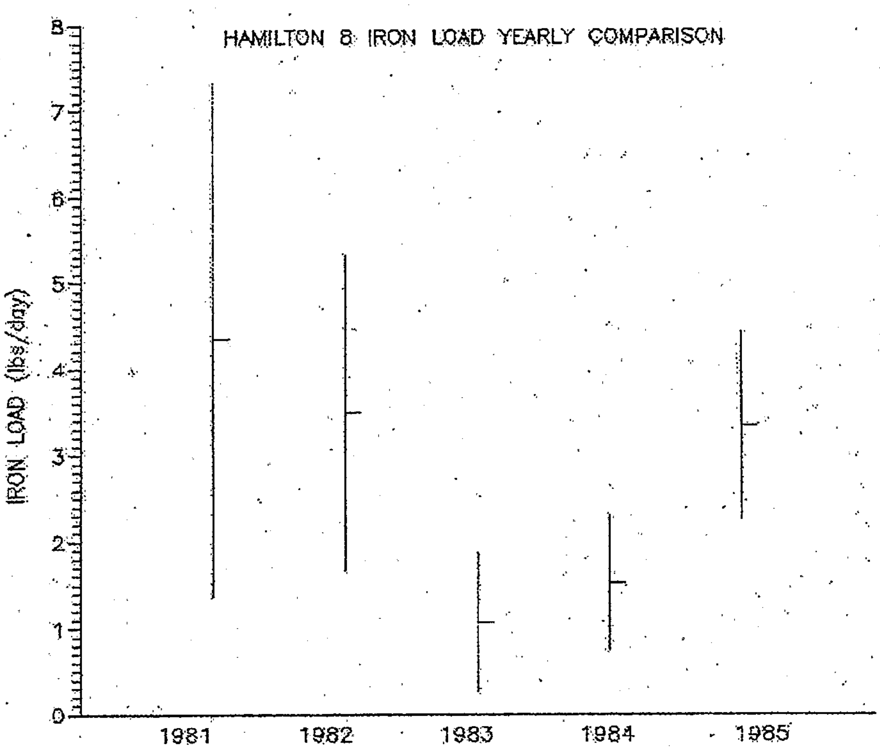


Figure 5.1y: Markson Acidity Load (1984-1986)

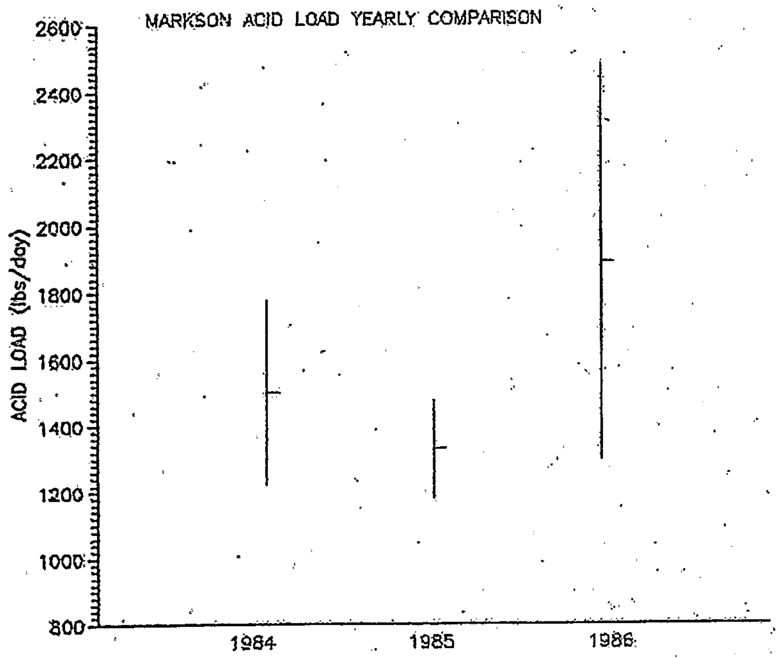


Figure 5.1z: Markson Iron Load (1984-1986)

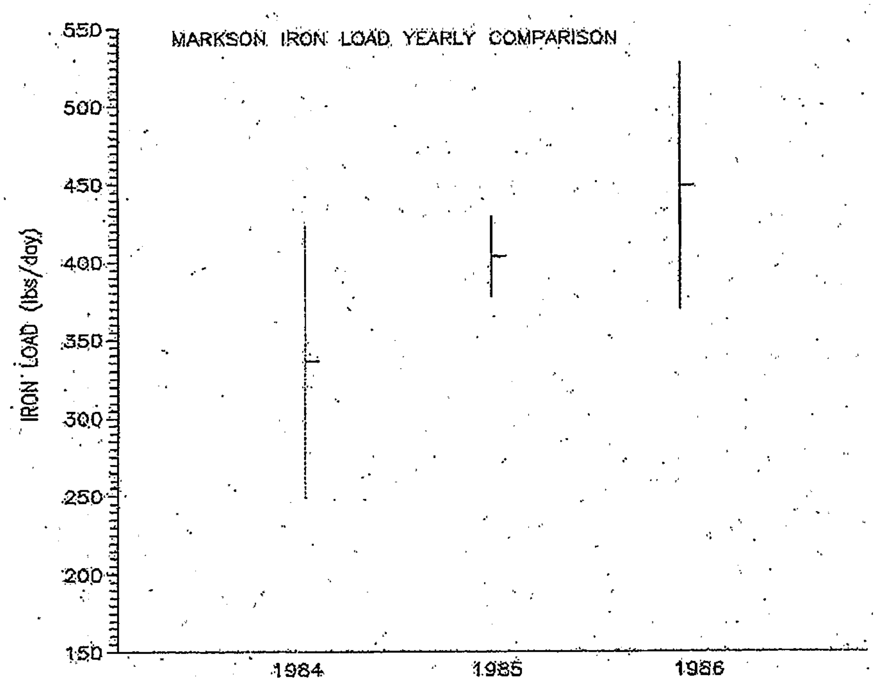


Figure 5.1aa: Acidity Load Data Comparison

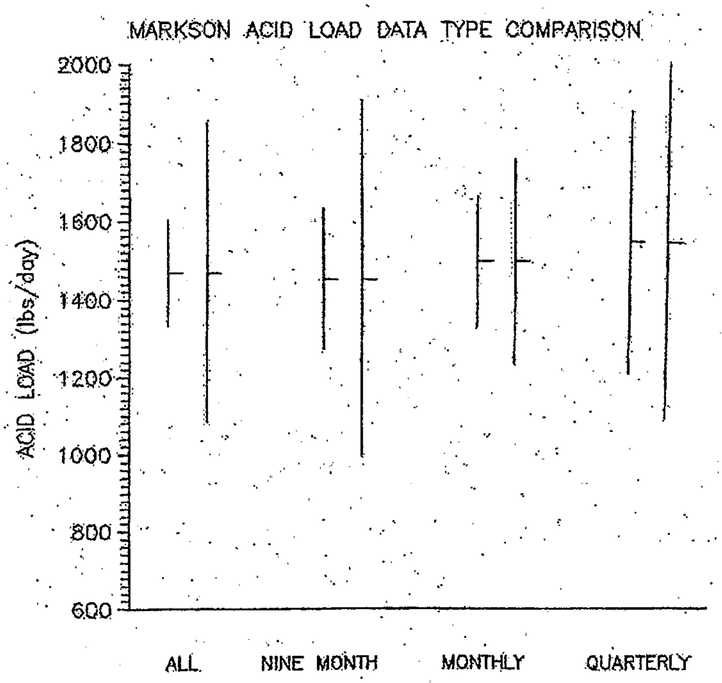
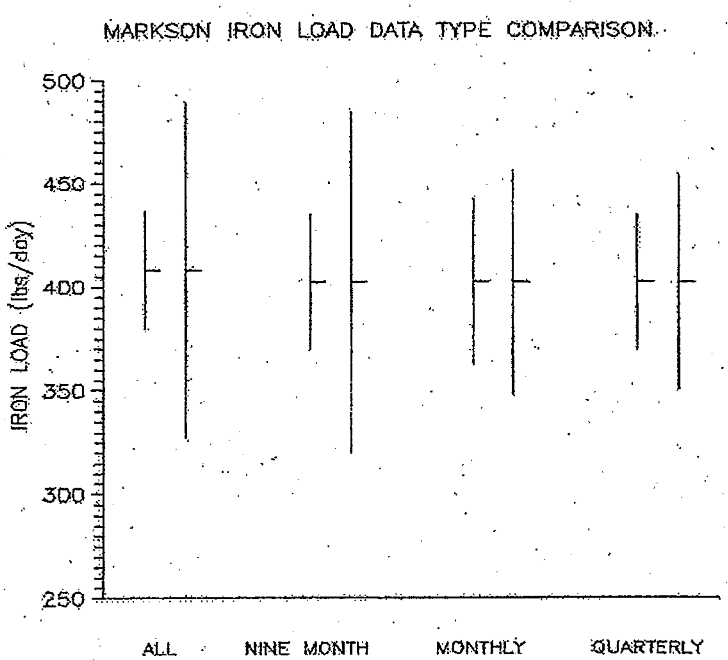


Figure 5.1ab: Markson Iron Load Data Comparison



5.1.2 Duration of Baseline Sampling

Previous study of these datasets (Griffiths, no date; 1987a-e; 1988a, b) observed that the months of November, December, and January typically exhibited behavior characteristic of median values and that extreme high and low flows and low and high concentrations were represented by the late winter/early spring and late summer/early fall months, respectively. This indicates the possibility that it may be acceptable to limit sample collection to a nine-month period that excludes the months of November, December, and January. To test this hypothesis, the long-term datasets were subsetted by eliminating all of the data from these three months, recalculating the baselines, and comparing the baseline median values for the full dataset and the nine-month subset. The results are shown in Table 5.1b.

Again using a 10 percent error criterion as a threshold for comparison, only two of the seven datasets (Clarion and Markson) showed less than 10 percent error in baseline median net acidity loads. Baseline iron loads showed similar results, with only two datasets (Fisher and Markson) showing less than 10 percent error. The average error for median acidity load was 10.75 percent. The average error for median iron load was 12.82 percent. The source of this error may be because even though the three excluded months typically have average flows, the median yearly flow may be greatly over or under estimated by excluding these months. For example, the median flow of the Arnot-3 discharge (Figure 5.1b) is much higher when using the nine-month data than with using the full 12-month dataset. Acidity loading rates, which are dominated by flows, parallel this effect (Figure 5.1d).

Based on this analysis, exclusion of the months of November, December, and January (in Pennsylvania and for areas with similar climates) poses a significant risk of not being representative of the entire water year and skewing the baseline loading rates, either higher or lower. Similarly, a sampling period of less than a full water year should be applied very cautiously before the results can be relied upon to develop a representative and statistically valid baseline.

5.1.3 Effects of Discharge Behavior on Baseline Sampling

Five of the seven discharges studied represent typical discharge behavior. These discharges exhibited relatively large seasonal fluctuations in flow rates with pollutant concentrations inversely proportional to flow. However, because changes in flow tend to be much greater than corresponding changes in concentration, flow tends to be the dominant factor in determining pollutant loading in these discharges. The result is a flow-dominated system with pollution loading rates that tend to closely follow the flow rate, although perhaps in a damped manner. This typical behavior is illustrated by the monthly flow and loading data from the Arnot-3 discharge (Figures 5.1e and 5.1f).

The remaining two discharges, Ernest and Markson, reacted very differently to changes in the baseline sampling period and interval. The Ernest discharge (a "slug response" discharge), yielded large percent errors for virtually every data subset. This discharge varied greatly in flow rate, concentration, and load, and responded very quickly to recharge events. These variations make representative monitoring very difficult. A baseline monitoring sampling interval that is too long (e.g., greater than monthly), can easily cause extreme events to be missed, or can over-represent extreme events if one happens to be sampled. Therefore, where this type of discharge behavior is evident, it would be prudent to use a shorter sampling interval (e.g., at least monthly) and/or expand the baseline sampling period.

The Markson discharge was the least affected by increasing the sample interval or using only nine months of data. Percent errors were relatively low regardless of the data subset used. This suggests that for discharges with typical steady-response behavior, it may be possible to obtain a suitable baseline using less frequent and possibly shorter sampling intervals. However, examination of the data on a year-by-year basis (Table 5.1c) indicates reason for caution. High volume discharges with very large storage reservoirs may be the most vulnerable to slow, long-term changes in flow caused by long-term or yearly variations in precipitation.

5.1.4 Year-to Year Variability

Annual median pollution loads and 95 percent confidence intervals for each of the seven discharges studied are presented in Table 5.1c. Although virtually all of the discharges showed some variability in confidence intervals from year-to-year, most of this variability was not statistically significant. There was only one discharge which exhibited statistically significant differences in baseline loading. The Ernest discharge, which has "slug response" behavior, tends to show extreme variability in both flow rate and load. As illustrated in Figure 5.1p, this was particularly the case in 1984, when the median acidity load was in excess of 5,000 lbs/day. This median is in contrast to all other years which had median acidity loads less than 2,000 lbs/day.

5.2 The Effects of Natural Seasonal Variations and Mining Induced Changes in Long-term Monitoring Data

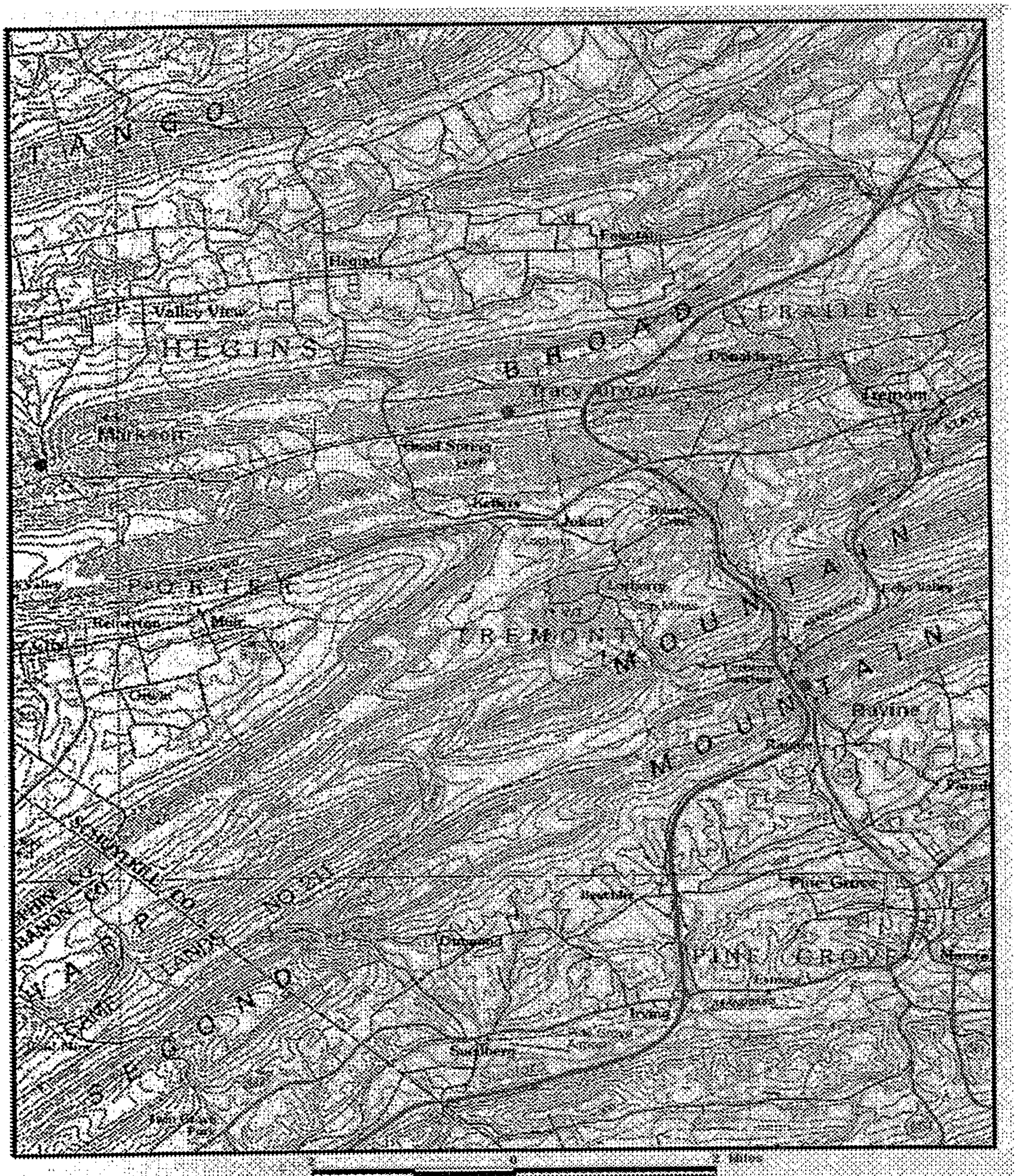
A primary reason for establishing a baseline pollution load prior to remining is to distinguish between natural seasonal variations and mining-induced changes in flow and water quality that may occur during remining and following reclamation. The reasons for using a sufficient number of samples, an adequate duration of sampling, and an acceptable sampling interval for establishing baseline pollution load are discussed throughout this document, and in EPA's Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits (USEPA, 2001; EPA-821-B-01-014).

The purpose of Section 5.2 is to: (1) depict the magnitude of natural seasonal variations of flow and water quality in several large abandoned underground mine discharges that were closely monitored for numerous years, and (2) provide examples of significant mining-induced changes in baseline pollution load at remining sites in Pennsylvania. Abandoned underground mine discharges (Markson, Tracy Airway, and Jeddo Tunnel) from the Pennsylvania Anthracite Coal Region are used to demonstrate the magnitude and patterns of natural seasonal variations. These discharges have been equipped with continuous flow recorders, and water quality analysis (at monthly or lesser sampling intervals) is available.

The Markson discharge is located approximately 1.2 miles (2 km) upstream from the Rausch Creek Treatment Plant operated by PA DEP and Schuylkill County (Figure 5.2a). This discharge emanates from an airway of an abandoned colliery at an elevation of 865 feet, and is a principal contributor to the acid load treated at the Rausch Creek Treatment Plant. The Tracy Airway discharge from another abandoned colliery is located 5.1 miles (8.3 km) east of the Markson discharge, and emanates from a mine-pool at an elevation of 1153 feet. The Tracy Airway discharge accounts for the largest iron load of all mine drainage discharges within the Swatara Creek watershed. The extensive data that were collected for both Markson and Tracy (Section 5.2.2) discharges is not typical of remining operations. These data were collected as the result of interest in diverting the discharges to a nearby treatment facility.

The U.S. Geological Survey (USGS) operates several gauging stations within the Swatara Creek Watershed as part of an EPA Section 319 National Monitoring Program (NMP) Project (the first of these projects in the United States focused on coal mine drainage problems) in cooperation with PA DEP, Schuylkill County, and other cooperators. The USGS station at Ravine shown in Figure 5.2a is the principal downstream gauge of the NMP project and is equipped with continuous flow and water quality recorders. The Markson discharge, Tracy discharge, and Ravine Station are located within a 5 mile radius, and therefore, should have been subjected to nearly equivalent amounts of precipitation, and duration and intensity of storm events during the period of record.

Figure 5.2a: Mine Discharge Map



5.2.1 Markson Discharge

The Markson discharge is characterized as a "steady response" type of discharge, where flow rate may vary seasonally, but changes in acidity concentrations or other water quality parameters are minimal or damped (Smith, 1988; Hornberger et al., 1990; and Brady, 1998). In a 1988 study of Markson data containing approximately 100 samples collected at weekly intervals from 1984 to 1986, Griffiths (1988a) found a lack of wide variation in all variables except flow, and found a lack of any strong relationship between pairs of variables (e.g., flow and acidity) except for an inverse correlation between iron and flow.

Monthly and annual variations in flow and concentrations of sulfate, acidity, pH, iron, and manganese are shown for an eight year period (1992-1999) in Figure 5.2b. The data were plotted on a logarithmic scale to demonstrate the range of variations in all of these variables on a single plot. Large annual variations in flow are apparent and appear to be inversely related to variations in sulfate and iron concentrations. Variations in acidity and manganese concentrations are more subtle, and do not readily show a strong relationship to flow variations.

Figure 5.2c depicts the relationships between the same flow and sulfate concentration data for the Markson discharge plotted on linear scales, while Figure 5.2d depicts the relationships between the flow and acid concentration on linear scales. Both Figures 5.2c and 5.2d show a generally strong inverse relationship between flow and pollutant concentration.

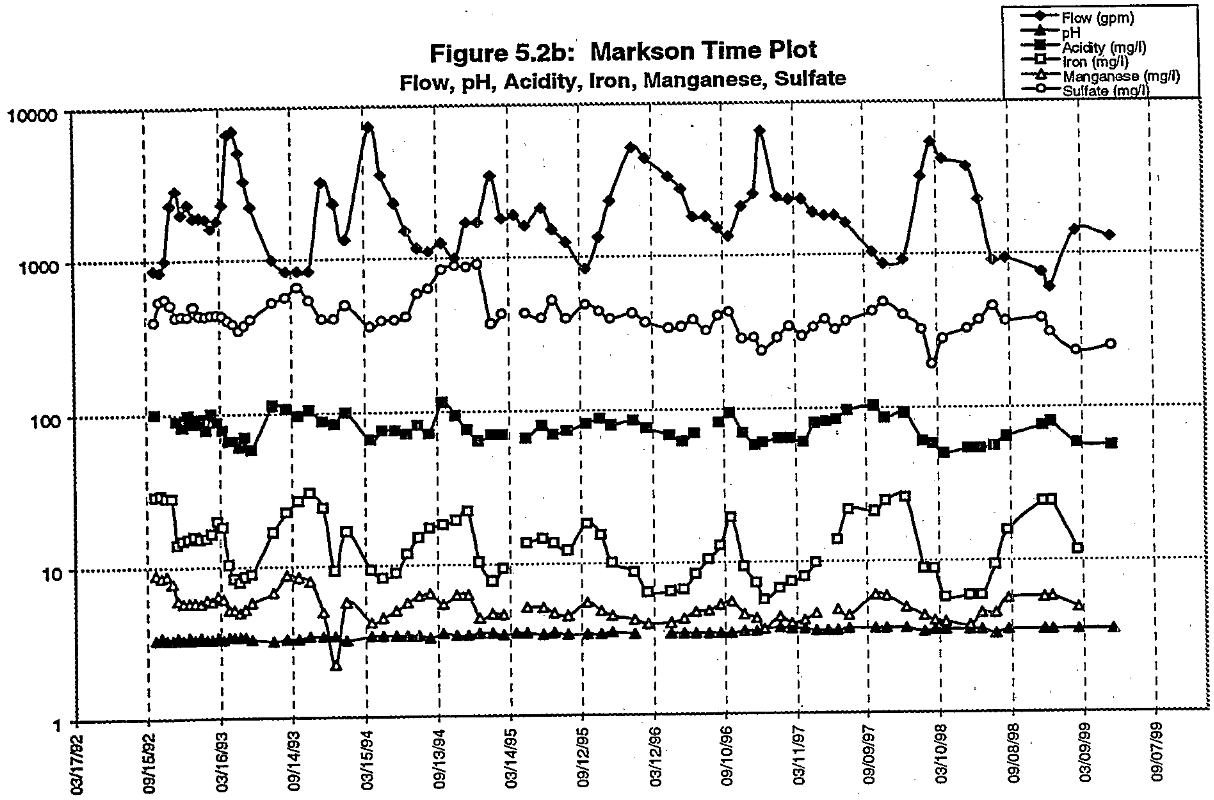


Figure 5.2b: Markson Time Plot
Flow, pH, Acidity, Iron, Manganese, Sulfate
Flow (gpm)
pH
Acidity (mg/l)
Iron (mg/l)
Manganese (mg/l)
Sulfate (mg/l)
10000
1000
100
10
1
03/17/92
09/15/92
03/16/93
09/14/93
03/15/94
09/13/94
03/14/95
09/12/95
03/12/96
09/10/96
03/11/97
09/09/97
03/10/98
09/08/98
03/09/99
09/07/99

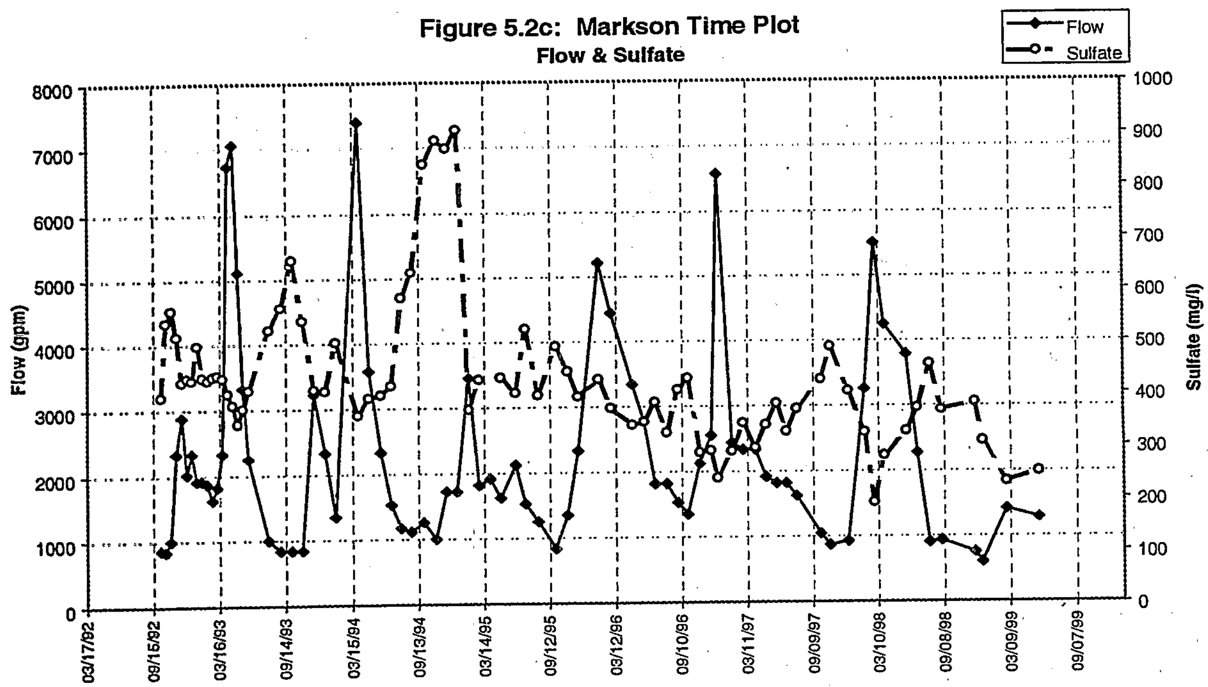


Figure 5.2c: Markson Time Plot
Flow & Sulfate
Flow
Sulfate
Flow (gpm)
Sulfate (mg/l)
8000
7000
6000
5000
4000
3000
2000
1000
0
1000
900
800
700
600
500
400
300
200
100
0
03/17/92
09/15/92
03/16/93
09/14/93
03/15/94
09/13/94
03/14/95
09/12/95
03/12/96
09/10/96
03/11/97
09/09/97
03/10/98
09/08/98
03/09/99
09/07/99

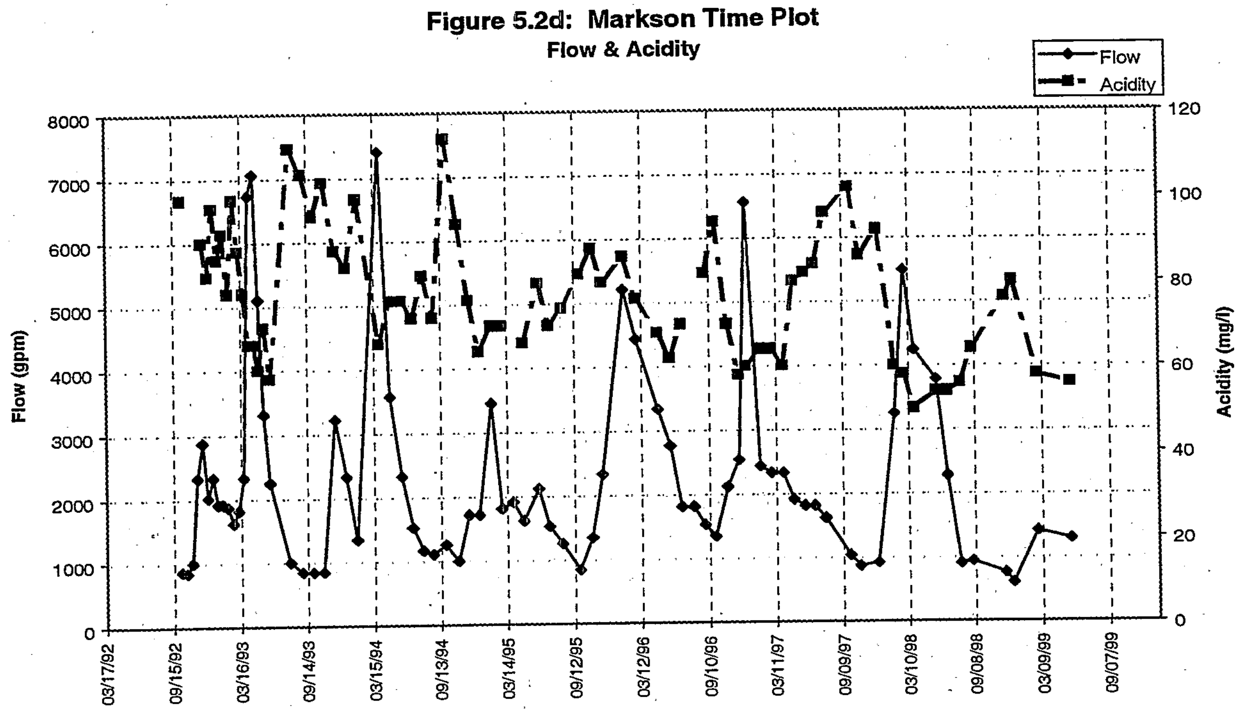


Figure 5.2e shows the relationship between monthly flow measurements and iron concentration. Figure 5.2f shows the corresponding relationships between flow and manganese concentration on linear scales. Both of these figures indicate a general inverse relationship between flow and metals concentrations in the Markson discharge. On a logarithmic scale (Figure 5.2b), manganese concentration did not appear to vary substantially in response to flow variations. This can be attributed to the relatively small range in manganese concentrations (2.2 to 8.9 mg/L) as compared to the range in iron concentrations (5.6 to 30.2 mg/L).

Figure 5.2e: Markson Time Plot
Flow & Iron

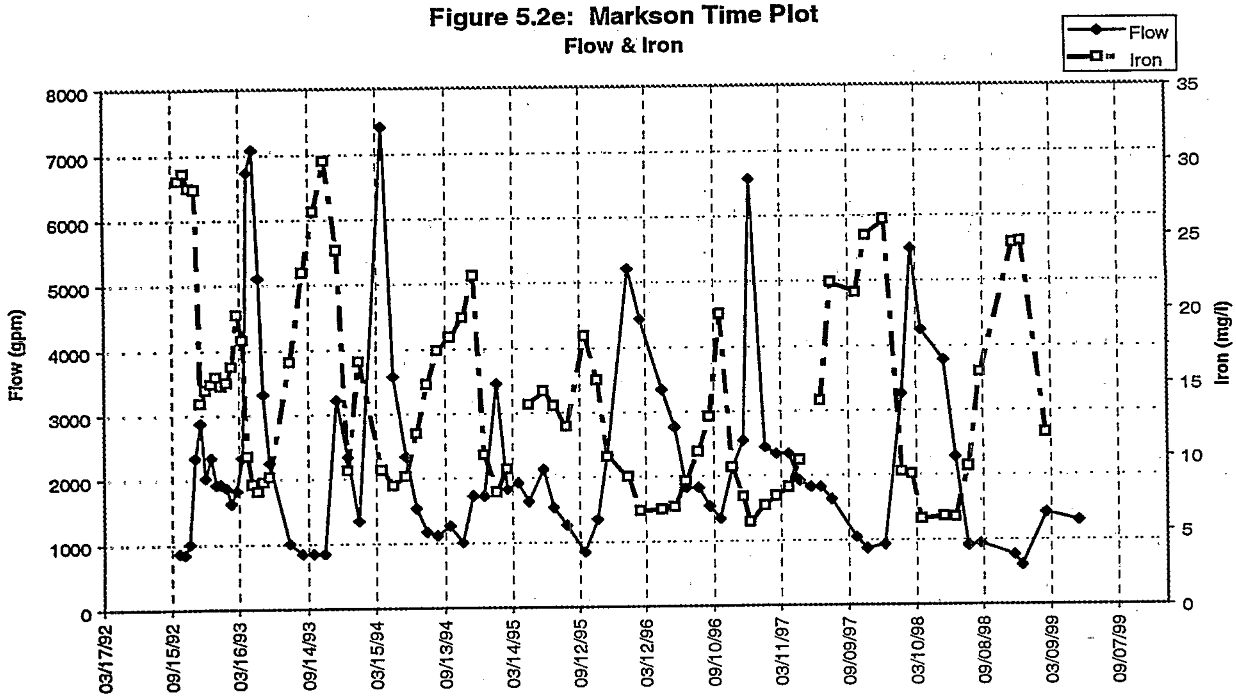
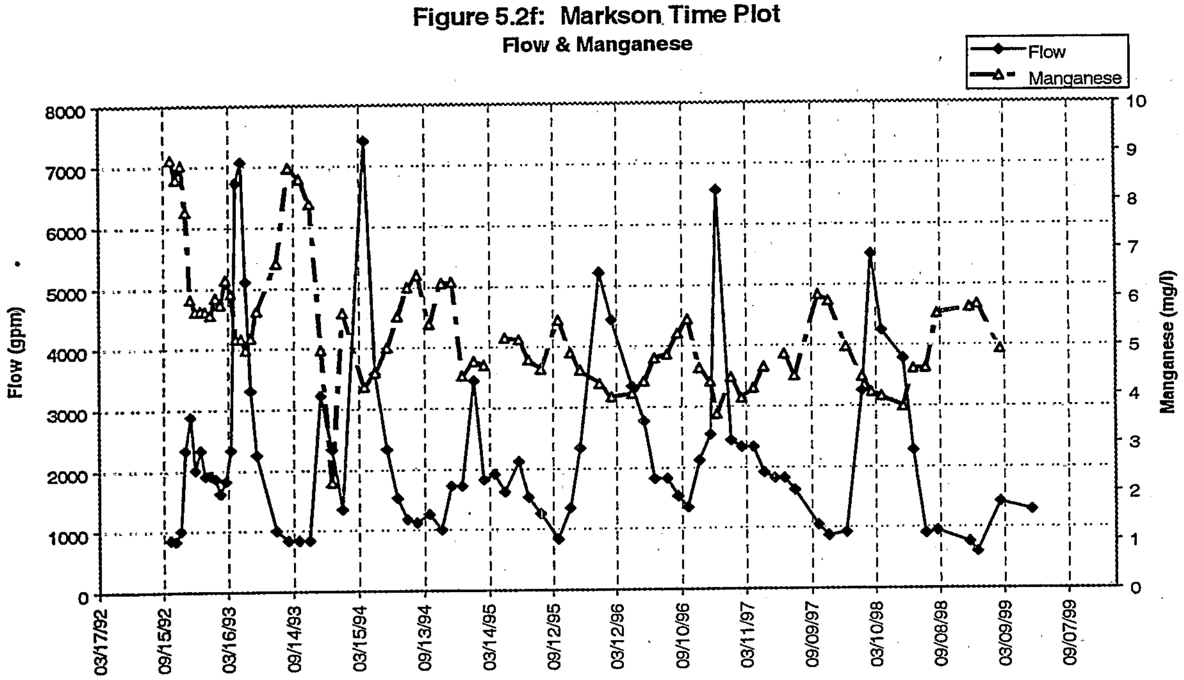


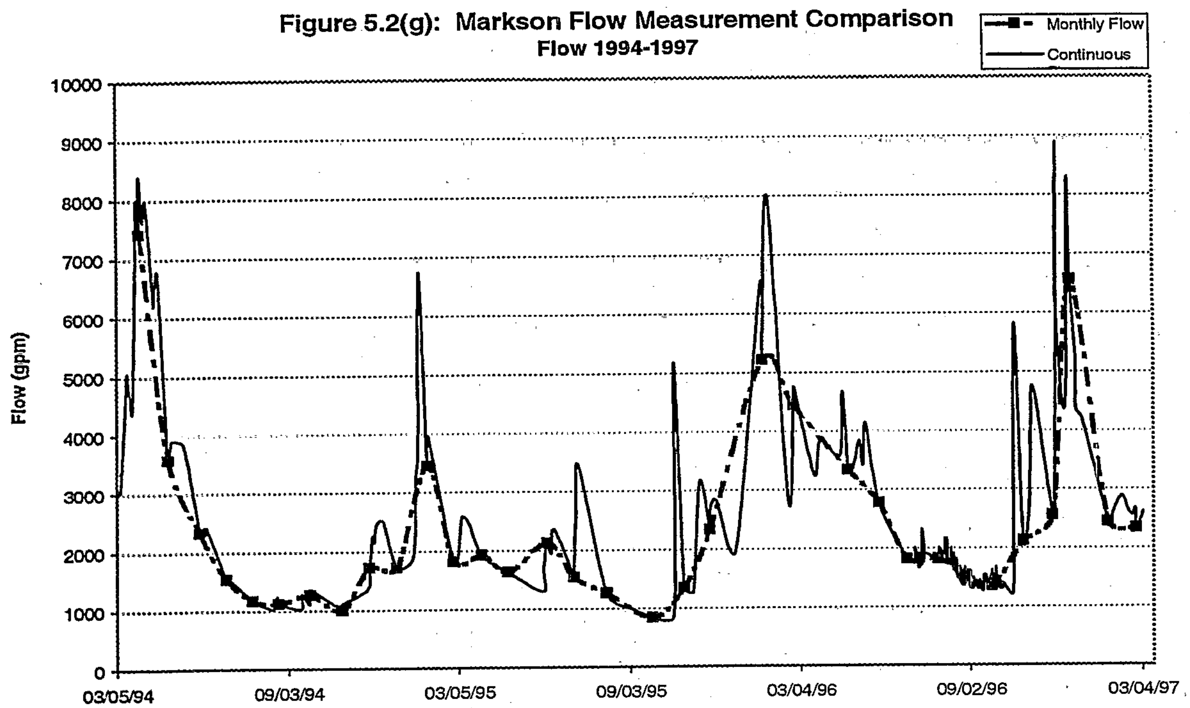
Figure 5.2f: Markson Time Plot
Flow & Manganese



In comparing the monthly measurements of flow for the eight year period shown in Figures 5.2b, 5.2c, and 5.2d, the following observations of annual variations can be made:

- There is a fairly regular annual pattern (the highest flow generally occurring in early to mid-March and the lowest flow generally occurring in mid to late September).
- Additional yearly peaks may occur (e.g., January 1996 and September 1999).
- The highest recorded monthly flow within water years can vary significantly (from a low of 3,500 gallons per minute in 1995 to a high of 7,500 gallons per minute in 1994).
- The lowest recorded monthly flow measurements are similar (ranging from 600 to 900 gallons per minute).
- The duration of high flow periods can vary substantially (e.g., 1996 compared to 1995).

The flow measurements presented in Figures 5.2b, 5.2c, and 5.2d represent the instantaneous flow recorded at the time monthly grab samples were collected for water quality analysis. Figure 5.2g shows the full range of continuous flow measurements for the three year period from March 1994 to March 1997, compared to the plot of the monthly data used in Figures 5.2b, 5.2c, and 5.2d. In compiling the continuous flow data, all of the continuous flow gauge recorder charts were evaluated to best define the extremes and duration of high and low flow events.



In comparing the continuous flow line to the instantaneous monthly flows, the following observations can be made:

- Continuous flow measurements exhibit much more variability than instantaneous monthly flow measurements.
- Monthly measurements missed some major storm events (July 1995, October 1995, October 1996, and November 1996).
- Although the highest annual continuous flow measurement usually corresponded to the same month as the highest annual monthly measurement, the differences between these measurements were very large (3,500 to 6,700 for 1995; 5,300 to 8,100 for 1996; and 6,600 to 8,900 for 1997).
- Monthly flow measurement may have occurred somewhat after the peak of a high flow event (February 1995) or somewhat before the peak (January 1996). This may explain most of the variations mentioned in the previous item.

- The differences between low flow events on the continuous flow plot and on the instantaneous monthly flow plots are relatively small for these three water years. This may imply that it is probably not difficult to define low flow periods with monthly samples.

5.2.2 Tracy Discharge

Monthly and annual variations in discharge flow and concentrations of sulfate, acidity, pH, iron, and manganese in the Tracy Airway discharge for the eight year period for 1992 through 1999 are shown in Figure 5.2h.

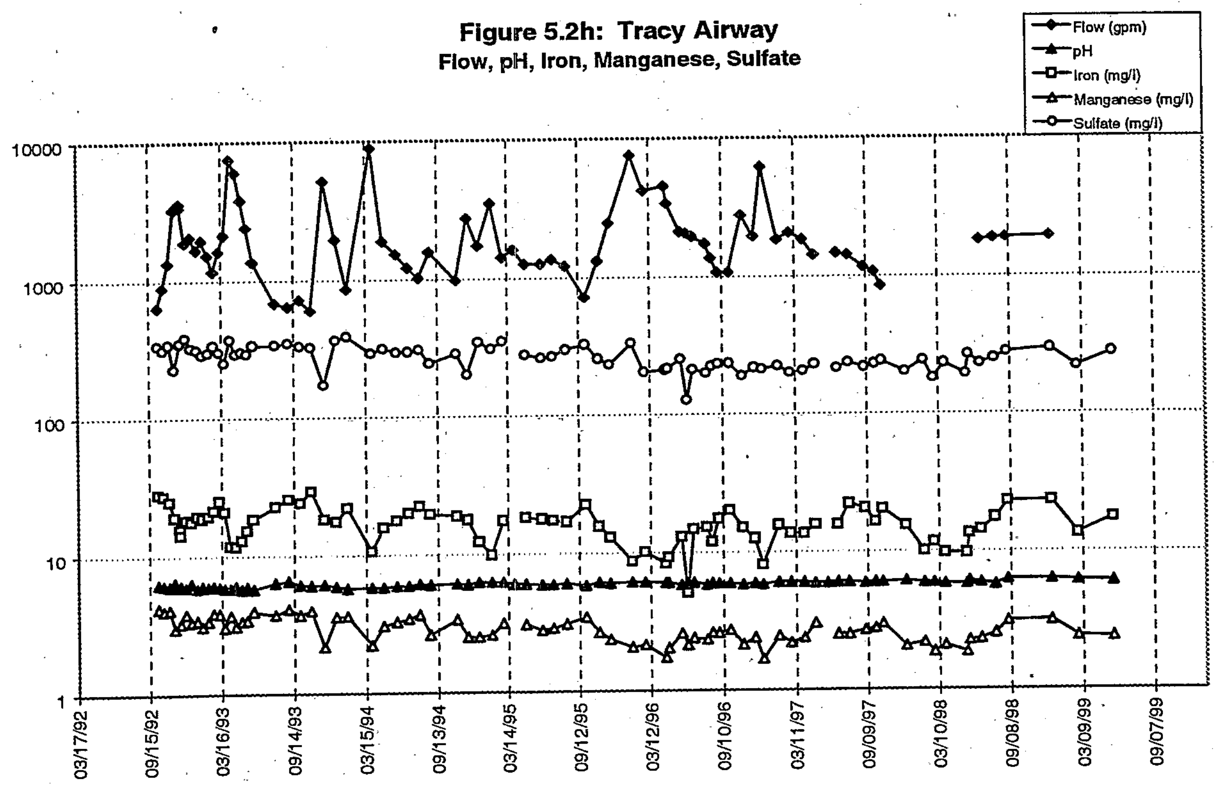


Figure 5.2h: Tracy Airway
Flow, pH, Iron, Manganese, Sulfate

The range and patterns of annual and long-term variations in flow and in concentrations of sulfate, iron, and manganese concentrations are similar to those for the same variables for the Markson discharge (Figure 5.2b). However, the water quality characteristics are fundamentally different in terms of pH, acidity, and alkalinity. The pH of the Tracy discharge ranged between 5.7 and 6.5 during the eight year period, and generally had an alkalinity concentration exceeding that of acidity. The pH of the Markson discharge ranged between 3.2 and 3.7 during the eight year period, had no net alkalinity (i.e., its pH is less than the titration end point), and generally had acidity concentrations around 100 mg/L. These distinct chemical differences in two discharges, emanating from similar mines in the same geologic structure and coal seam (the Donaldson Syncline), are attributable to stratification of large and deep anthracite minepools. The Tracy discharge is a "top-water" discharge from a relatively shallow groundwater flow system (at an elevation of 1153 feet), while the Markson discharge emanates from "bottom water" at a much lower elevation in the minepool (865 feet). The chemistry of stratified anthracite mine-pools is described by Brady et al. (1998) and Barnes et al. (1964). However, these discharges are similar in the relationship of flow and water quality to natural seasonal variations.

The monthly flow pattern of the Tracy discharge (Figure 5.2h) is very similar to that of the Markson discharge (Figure 5.2b), except the Tracy discharge flows appear to be somewhat more variable or peaked. When the annual patterns of high and low flows are compared, the Tracy discharge has two flow peaks (November 1995 and October 1996) that do not occur for the Markson discharge. These two peaks indicate storm events undetected by monthly sampling. The plot of continuous and monthly flow records for the Tracy discharge (Figure 5.2i) reveals that the Tracy discharge was sampled a short time before the November 1995 flow peak and well after the flow peak for October 1996 (continuous flow peak equals 6700 gpm, monthly flow equals 2700 gpm).

Figure 5.2i: Tracy Airway
Flow 1994-97

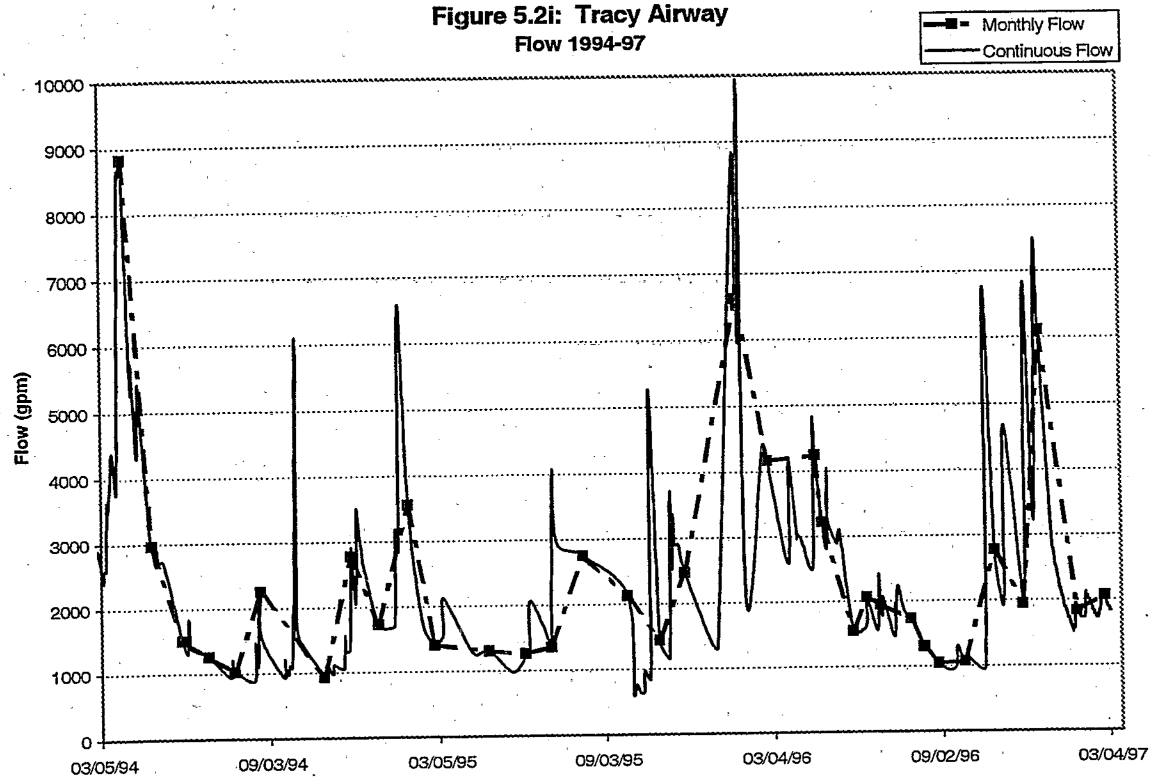
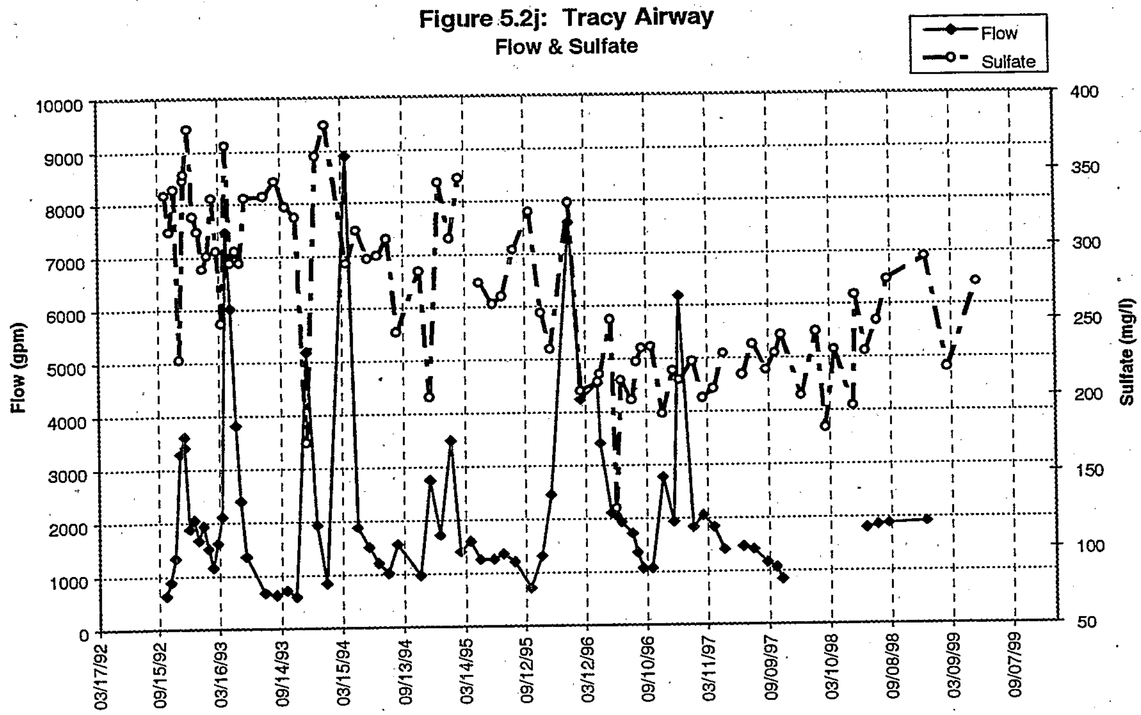
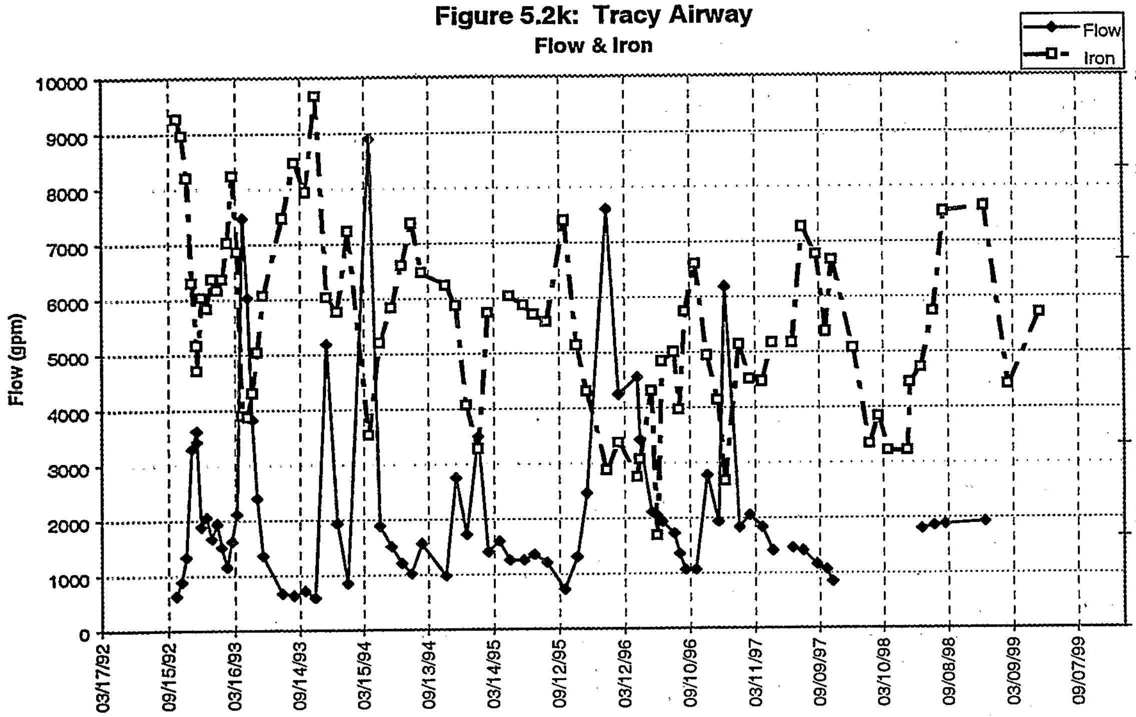
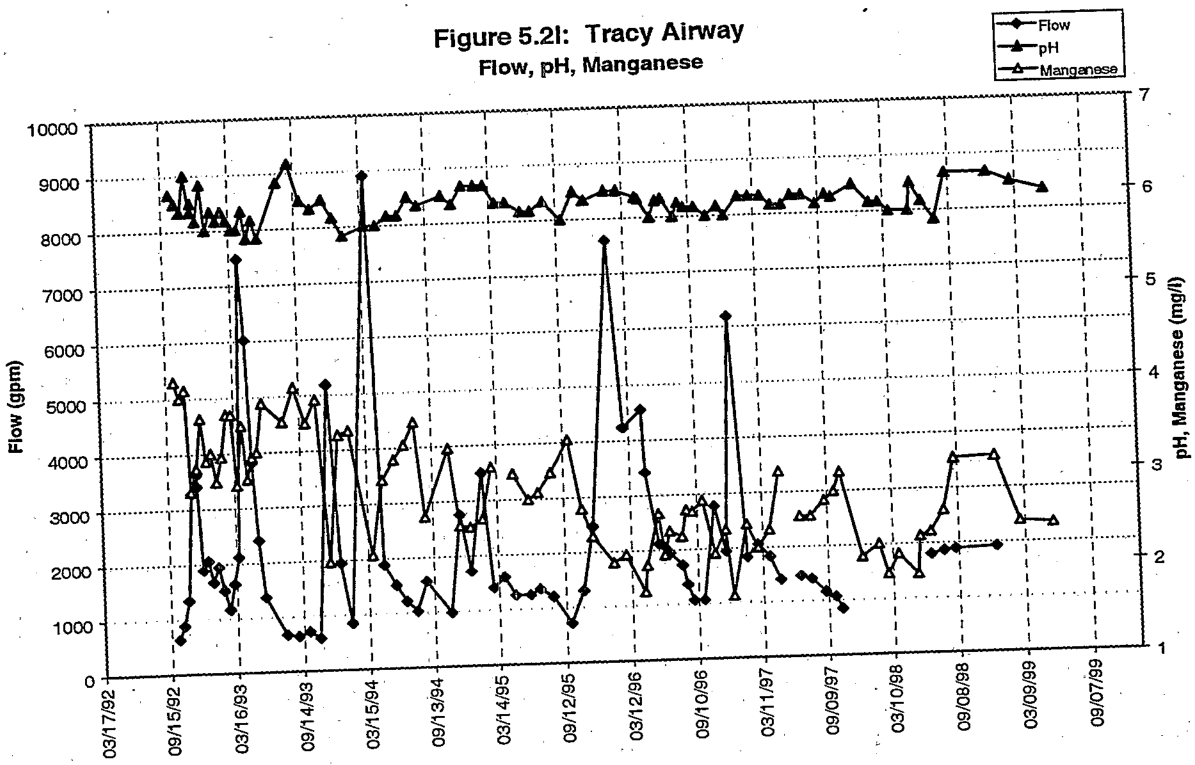


Figure 5.2j: Tracy Airway
Flow & Sulfate





In comparing continuous flow data (Figure 5.2i) to monthly flow data, most of the Markson discharge results were also apparent in the Tracy discharge. While the monthly samples correspond well during the first two major high flow events in 1994 (April and September), several major storm events go undetected in the succeeding data (e.g., October and November 1994, November 1995, and November 1996). In addition, there are several major storm events where the monthly sample was collected after (February, July, and August 1995, and October 1996) or before (January 1996) the peak recorded by continuous monitoring. The differences between the monthly and continuous recorder data peaks are most significant in February 1995 (3700 and 6700 gpm), January 1996 (6500 and 9900 gpm), and October 1996 (2700 and 6700 gpm). One interesting characteristic of the Tracy flow data is that the continuous flow monitor results for 1995 "bottoms out" several hundred gallons per minute below the monthly data, but corresponds well with the low flow continuous recorder data for other water years.



A strong inverse relationship between flow and pollutant concentration in the Tracy discharge is shown in Figures 5.2j, 5.2k, and 5.2l, with the highest flows corresponding to the lowest concentrations and the lowest flows corresponding to the highest concentrations for sulfate (Figure 5.2j), iron (Figure 5.2k), and manganese (Figure 5.2l). There appears to be a trend over time, where the range of median values for sulfate and manganese are diminished from 1992 through 1997.

Monthly flow and water quality relationships of the Markson and Tracy discharges, throughout the eight year period shown in Figures 5.2b through 5.2l, indicate a general inverse relationship between flow and concentration, but also show that the distribution, magnitude, and duration of high flow events is not uniform from water year to water year. In fact, sometimes the highest flow events appear during what is traditionally the low flow period of the water year (e.g., October 1996 and September 1999). These data suggest that a sampling interval length of not greater than one month, and a sampling duration of at least a water year (12 monthly samples) are necessary to document baseline flow and water quality variations, particularly if high flow events are important in establishing the baseline pollution load. The monthly and continuous flow data for the Markson and Tracy discharges (Figures 5.2g and 5.2i) show that representative sampling of storm events can be tricky, as isolated sampling events may not always capture the range and pattern of natural seasonal variations. This problem is illustrated in Figures 5.2g and 5.2i, where monthly flow measurements indicate high flows that are much lower than those measured by the continuous flow monitors, or where significant high-flow periods detected by continuous monitoring were undetected by the monthly measurements.

5.2.3 Swatara Creek Monitoring Station

USGS has been sampling water quality and flow characteristics of Swatara Creek in Schuylkill and Lebanon Counties, Pennsylvania since before 1960. The results of this data collection are found in numerous publications including McCarren et al. (1961) and Fishel and Richardson (1986). The USGS Ravine Station shown in Figure 5.2a has been a key station because it is located on the main stem of Swatara Creek immediately below the confluence of several tributaries draining the coal operations in the Swatara Creek headwaters. Below the Ravine Station, the Swatara Creek watershed changes to a more agricultural land use without acid mine drainage contributions to water quality. Figures 5.2m, 5.2n, and 5.2o contain a series of plots of the storm-flow hydrograph and continuous measurements of specific conductance and pH for a five day period in December 1996 (Cravotta, personal communication). These figures also show water quality data for sulfate, suspended solids, and iron (total and dissolved) that were collected by automatic samplers for flows resulting from this storm flow period. These figures indicate that water quality data peaks for suspended solids and iron precede the peak for flow. According to Cravotta (1999), the occurrence of these concentration peaks prior to peak flow are the result of scour and transport of stream bed deposits.

Figure 5.2m: Swatara Creek Flow and Sulfate Data

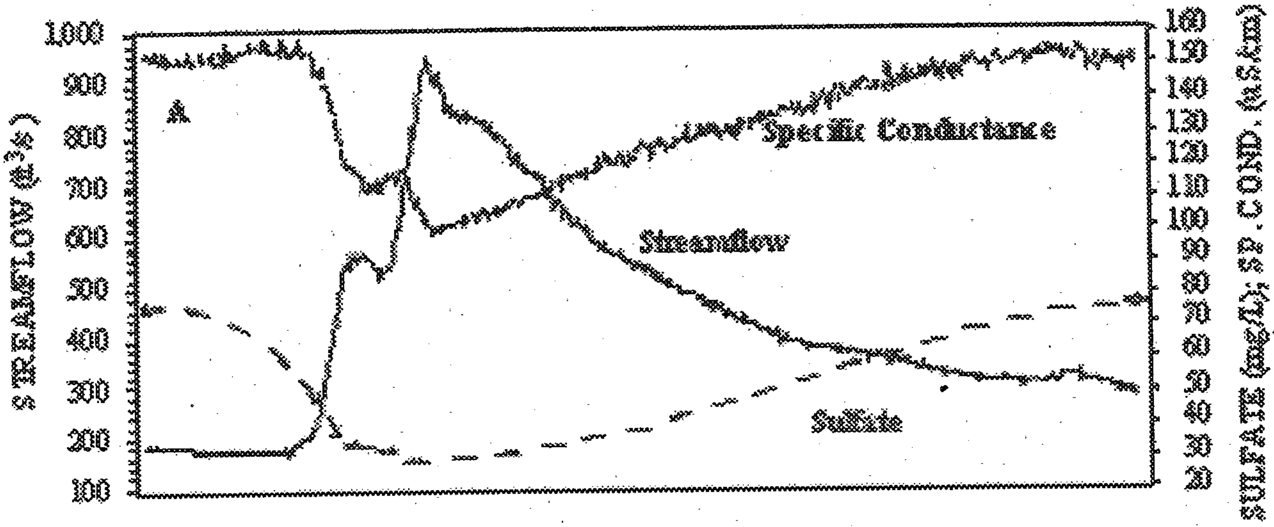


Figure 5.2n: Swatara Creek Flow and Suspended Solids Data

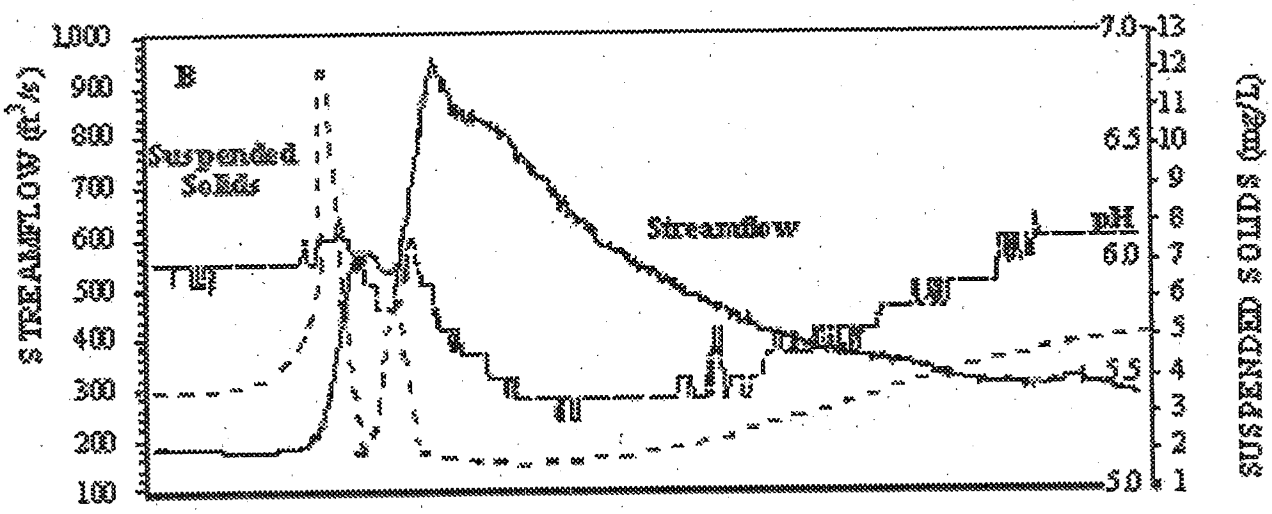
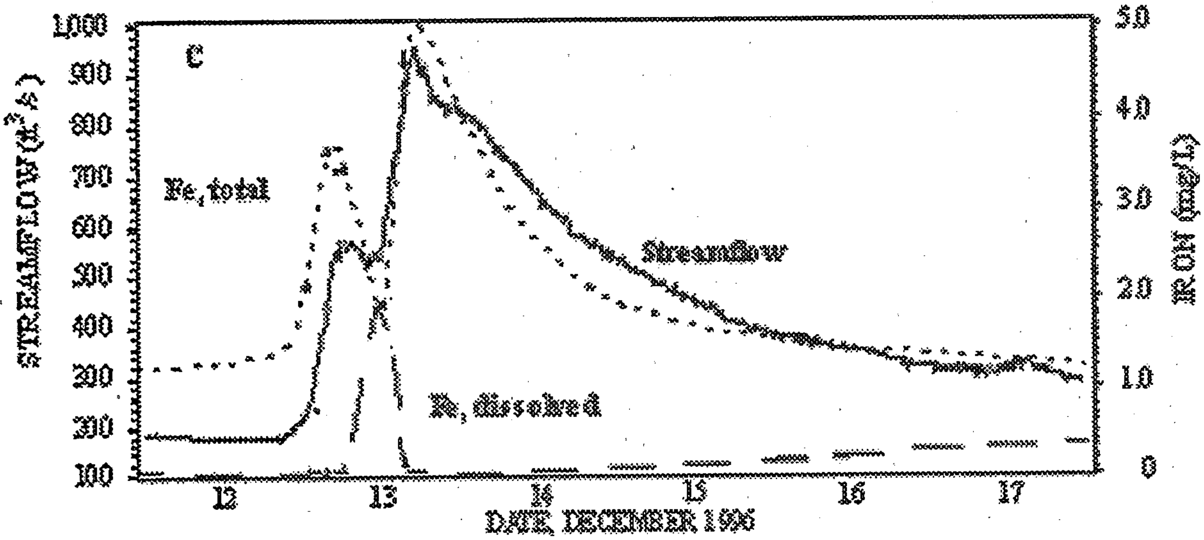


Figure 5.2o: Swatara Creek Flow and Iron Data



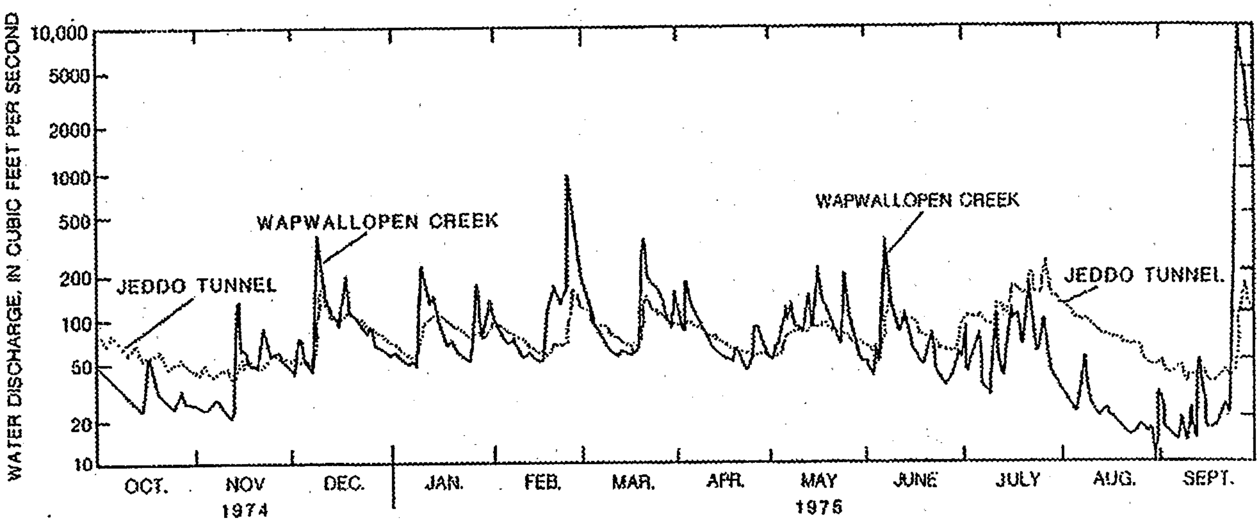
5.2.4 Jeddo Tunnel Discharge

The Jeddo Tunnel mine discharge near Hazleton Pennsylvania is the largest abandoned underground mine discharge in the Eastern Middle Field of the Anthracite Region, and is among the largest mine drainage discharges in Pennsylvania. The Jeddo Tunnel has a total drainage area of 32.24 square miles, and its underground drainage system collects and discharges more than half of the precipitation received in the drainage area (Balleron et al., 1999). The flow of this discharge was monitored with a continuous recorder from December 1973 through September 1979 by the USGS in cooperation with Pennsylvania Department of Environmental Resources.

The results of that monitoring for the water year from October 1, 1974 through September 30, 1975 are shown in Figure 5.2p (Growitz et al., 1985). During that year, the discharge ranged from 36 to 230 cfs (16,157 to 103,224 gpm). The Jeddo Tunnel discharge flows are compared to the stream-flow of Wapwallopen Creek (approximately 10 miles north of the Jeddo Tunnel).

The Wapwallopen Creek drains an area of 43.8 square miles and has a measured mean discharge of 78 cfs (35,008 gpm) (Growitz et al., 1985). Growitz et al. found that the response of the Jeddo Tunnel discharge to precipitation events is considerably less than that of the Wapwallopen Creek, and that during large storm events, the Jeddo Tunnel data peaked later than the stream discharge.

Figure 5.2p: Jeddo Tunnel Discharge and Wapwallopen Creek Flow Data



The continuous flow recording station at the mouth of the Jeddo Tunnel was reconstructed and operated by USGS from October 1995 through September 1998 in cooperation with PA DEP, the Susquehanna River Basin Commission, US EPA, and other project cooperators. Figure 5.2q (from Balleron et al., 1999) shows variations in the flow of this discharge during this period. The average annual discharge flow was 79.4 cfs (35,635 gpm) and the range of recorded flow measurements was between 20 cfs (8,976 gpm) in October 1995 and 482 cfs (216,322 gpm) in November 1996, following 3.89 inches of rainfall (Balleron, 1999). Figure 5.2r shows a graph of precipitation data from Hazleton Pennsylvania for the period from October 1995 through September 1998. This graph was plotted from data contained in Balleron (1999).

Figure 5.2q: Jeddo Tunnel Flow Data

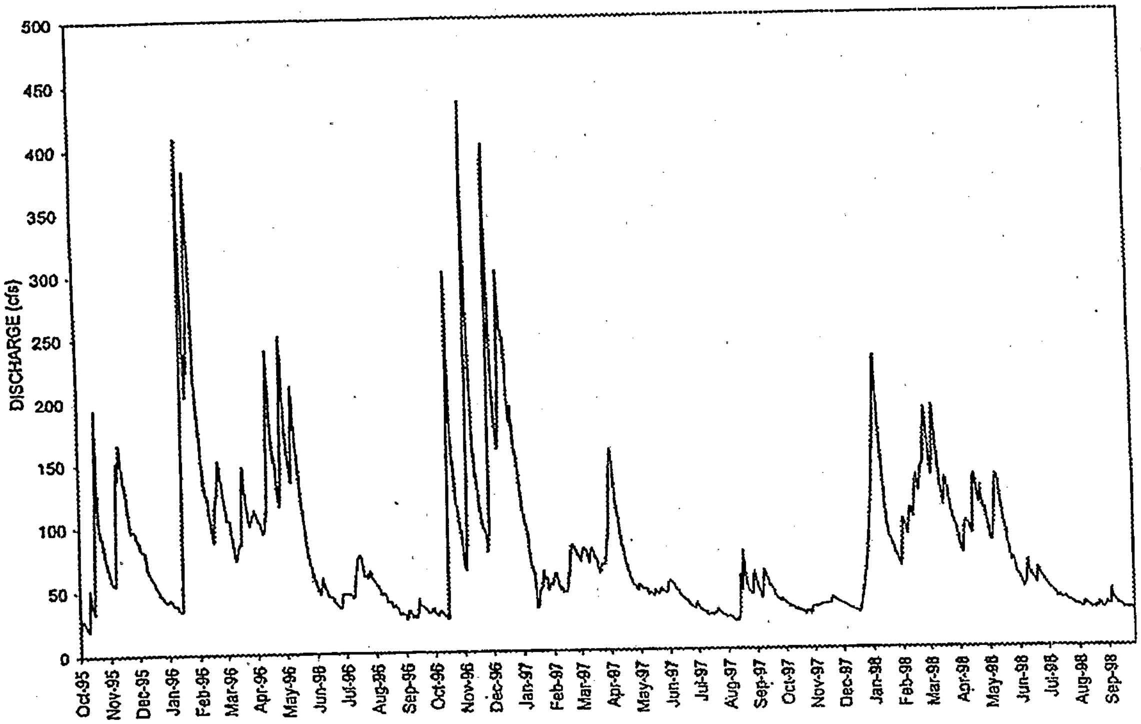
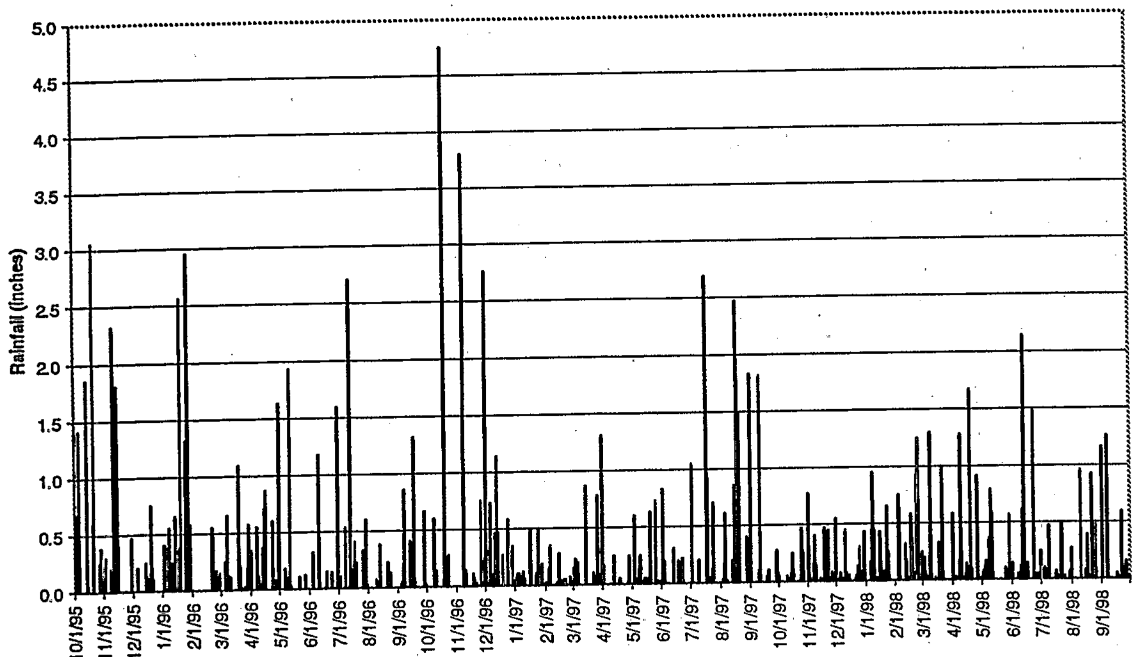


Figure 5.2(r): Precipitation Data From Hazleton, Pa



5.3 Case Studies

Baseline pollution loading in pre-existing discharges must be measured and monitored accurately to determine to what extent polluting conditions are affected by remining operations. The Fisher, McWreath, and Trees Mills remining sites in the Pennsylvania Bituminous Coal Region are presented as case studies in Section 5.3 to demonstrate significant changes in flow, water quality, and pollution load resulting from remining and reclamation activities. The case studies also demonstrate how a regular monitoring program can be used to evaluate and document both pre- and post-remining water quality from a pre-existing discharge. In each of these cases, monitoring was conducted at monthly intervals and proved to be adequate to document baseline conditions and to demonstrate post-remining changes in water quality. These case studies also illustrate the water quality and quantity changes that are typical of remining operations

5.3.1 Fisher Remining Site

The Fisher remining site is located in Lycoming County, Pennsylvania. Prior to remining, the surface of the site was extensively disturbed by abandoned surface mine pits and spoil piles. A large abandoned underground mine known as the Fisher deep mine occupied much of the sub-surface. The principal discharge (monitoring point M-1) was the main concern during the remining permit process. Baseline pollution load data collection took place between 1982 and 1985. The original remining permit was issued on November 5, 1985, and remining operations commenced by February 1986. Final coal removal occurred in June 1995 and backfilling was completed within the permit area by February 1996.

The best management practices employed on the Fisher remining site include: (1) daylighting the abandoned Fisher deep mine, (2) regrading abandoned spoils and backfilling abandoned pits, (3) alkaline addition, and (4) biosolids used for revegetation enhancement. Alkaline addition was accomplished with 140,000 tons of limestone fines on the two most recently permitted areas, resulting in an alkaline addition rate of approximately 400 tons per acre over an area of approximately 350 acres. Biosolids were applied to approximately 500 acres.

The relationships between the sulfate, iron, and manganese concentrations, and flow in the M-1 discharge for the period between 1982 and 1999, are shown in Figure 5.3a. There are several trends in the relationships resulting from remining activities. While iron concentrations decreased over time, sulfate and manganese concentrations increased. Discharge flow increased following backfilling (1996), probably because this point became the down gradient drain for greater than 500 acres of unconsolidated mine spoil aquifer materials. The most significant change in pollutant concentration was in net acidity (Figure 5.3b). Prior to activation of the remining permit, the acidity concentration was typically in the range of 100 to 200 mg/L. The effect of remining was to turn a distinctly acidic discharge into one that is now distinctly alkaline (i.e., post-mining net acidity concentrations of 0 through –75 mg/L).

Figure 5.3a: Fisher Mining MP1
Flow, Iron, Manganese, Sulfate

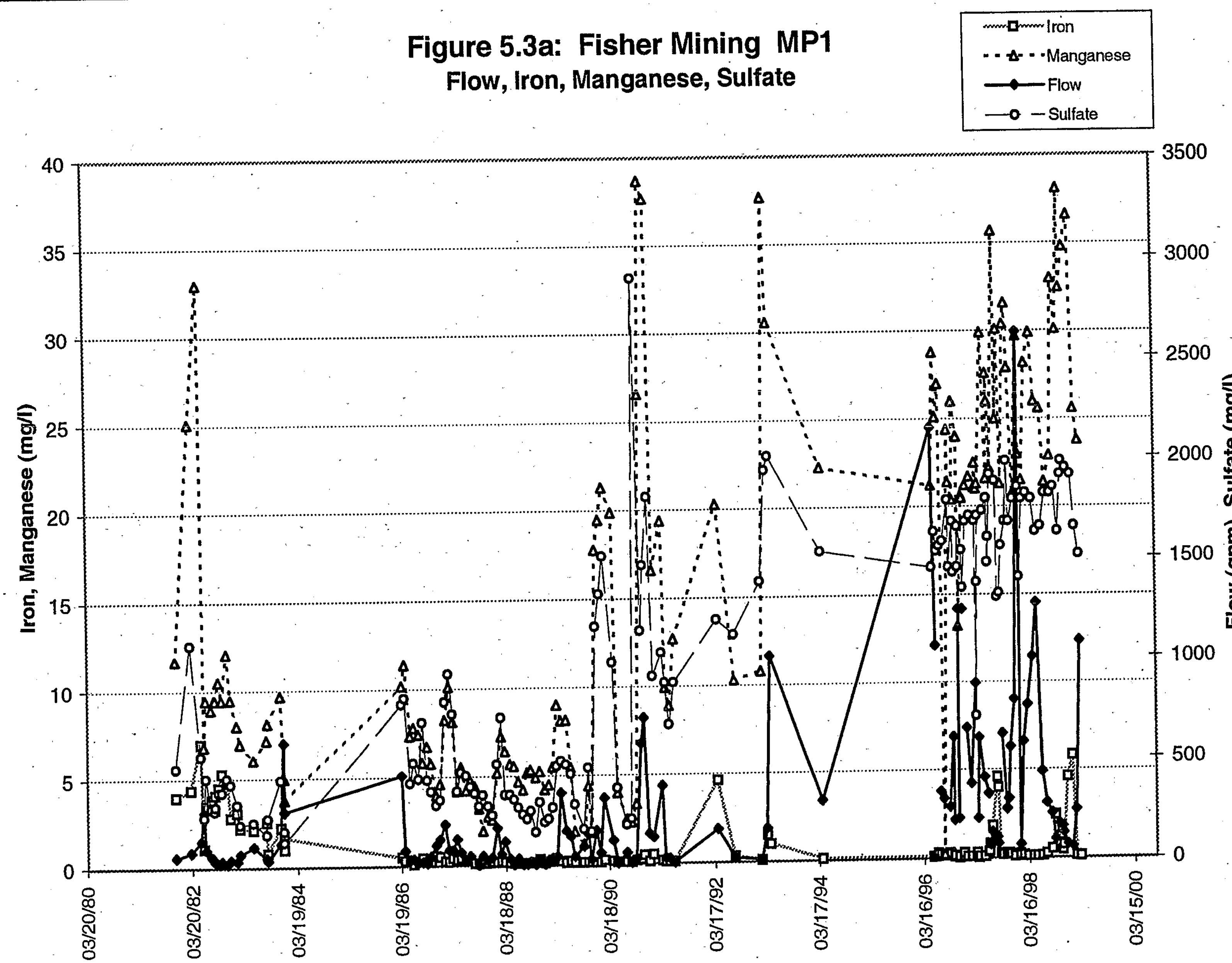


Figure 5.3b: Fisher Mining MP1
Net Acidity

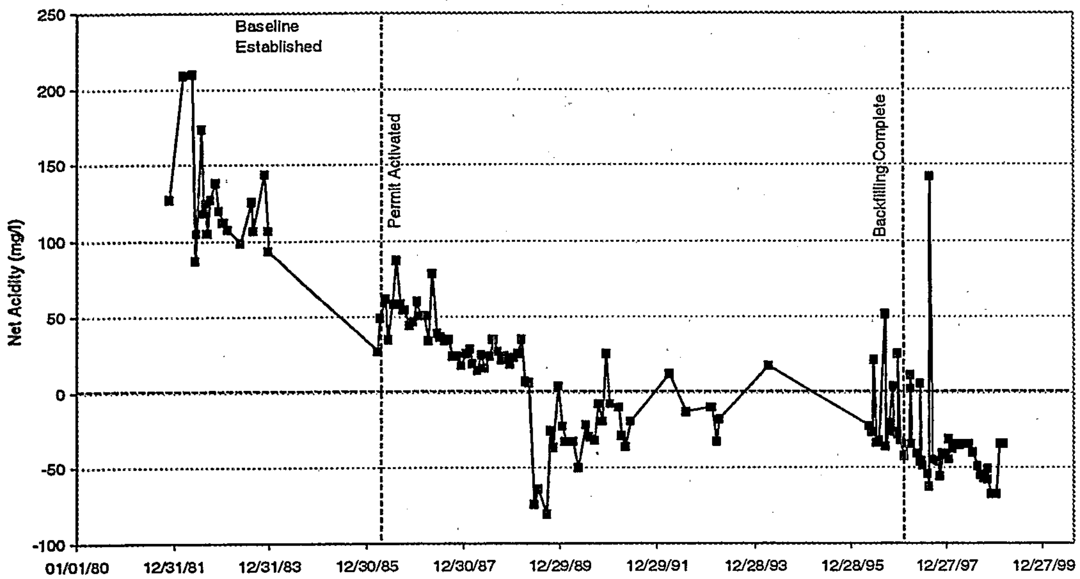
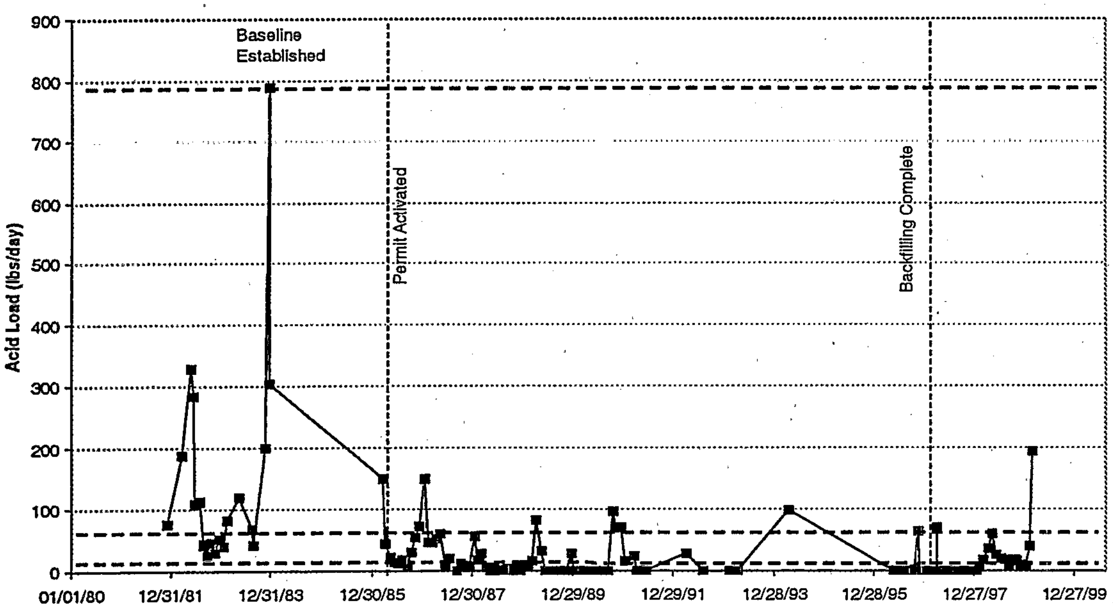


Figure 5.3c: Fisher Mining MP1
Acid Load



The Fisher permit (and most remining permits in Pennsylvania) was written in a format that evaluates remining performance on a pollution load basis rather than a concentration basis. As part of the baseline pollution load computation, quality control limits are based upon the approximate 95 percent tolerance limits around the frequency distribution, and the median is used as the measure of central tendency of the frequency distribution (see Section 1.0, Table 1.2a). Acidity loading prior to permit activation (baseline establishment), during open pit mining, and following backfilling and reclamation for the M-1 discharge are shown in Figure 5.3c. The upper and lower of the three horizontal dashed lines correspond to the upper and lower 95 percent tolerance limits for the pre-mining baseline pollution load. The middle dashed line is the baseline median acidity load (67.9 pounds per day). The median acidity load for the three year period following backfilling (1996 through 1999) is 0 pounds per day, showing improvement in water quality. Figure 5.3d shows corresponding improvement in the iron load (pre-mining baseline median iron load was 1.36 pounds per day; median of the three years following backfilling is 1.04 pounds per day). The differences in iron load and in net alkalinity concentration during these time periods are presented in Figures 5.3e and 5.3f respectively.

Figure 5.3(d): Fisher Mining MP1
Iron Load

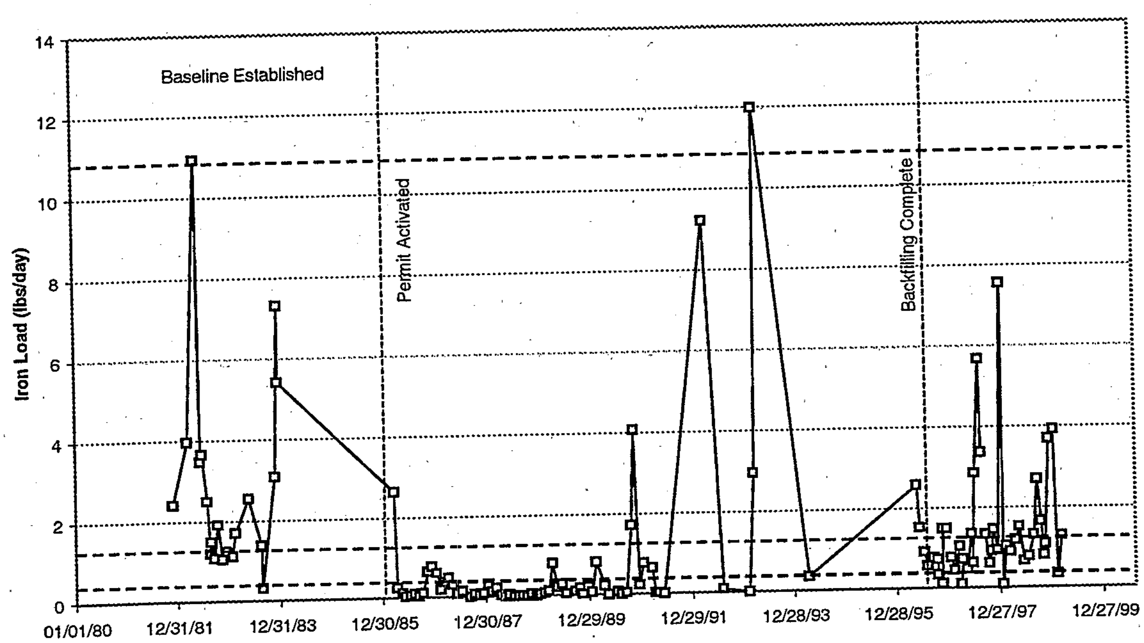


Figure 5.3e: Iron Load Boxplot

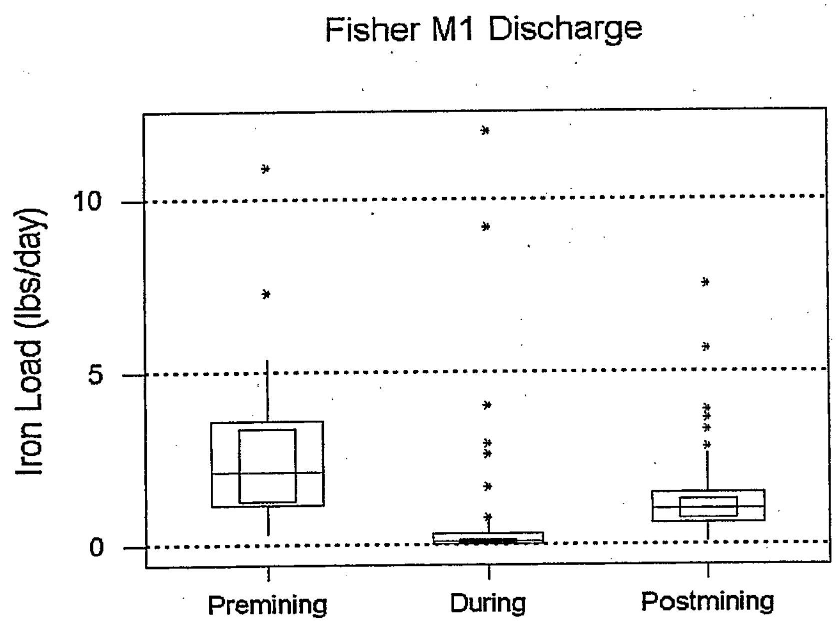
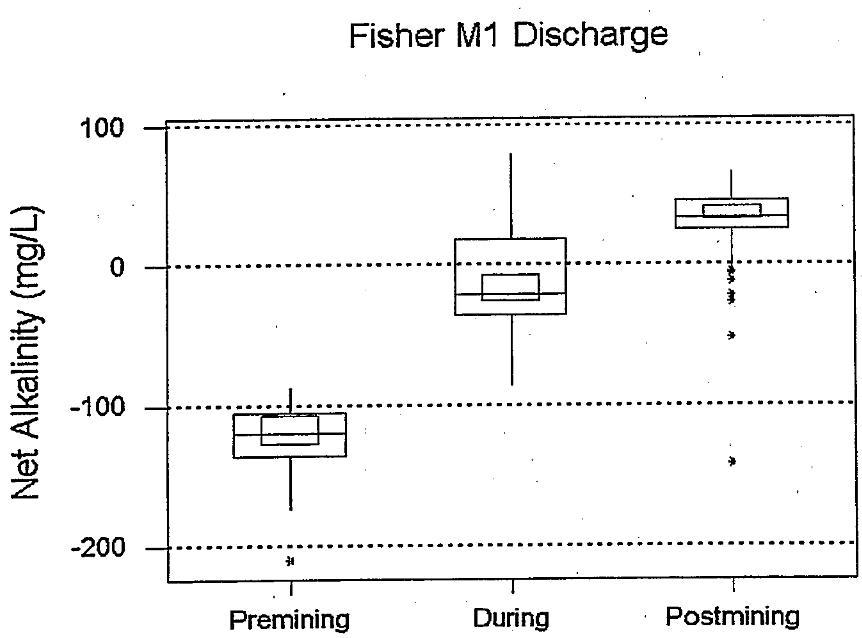


Figure 5.3f: Net Alkalinity Boxplot

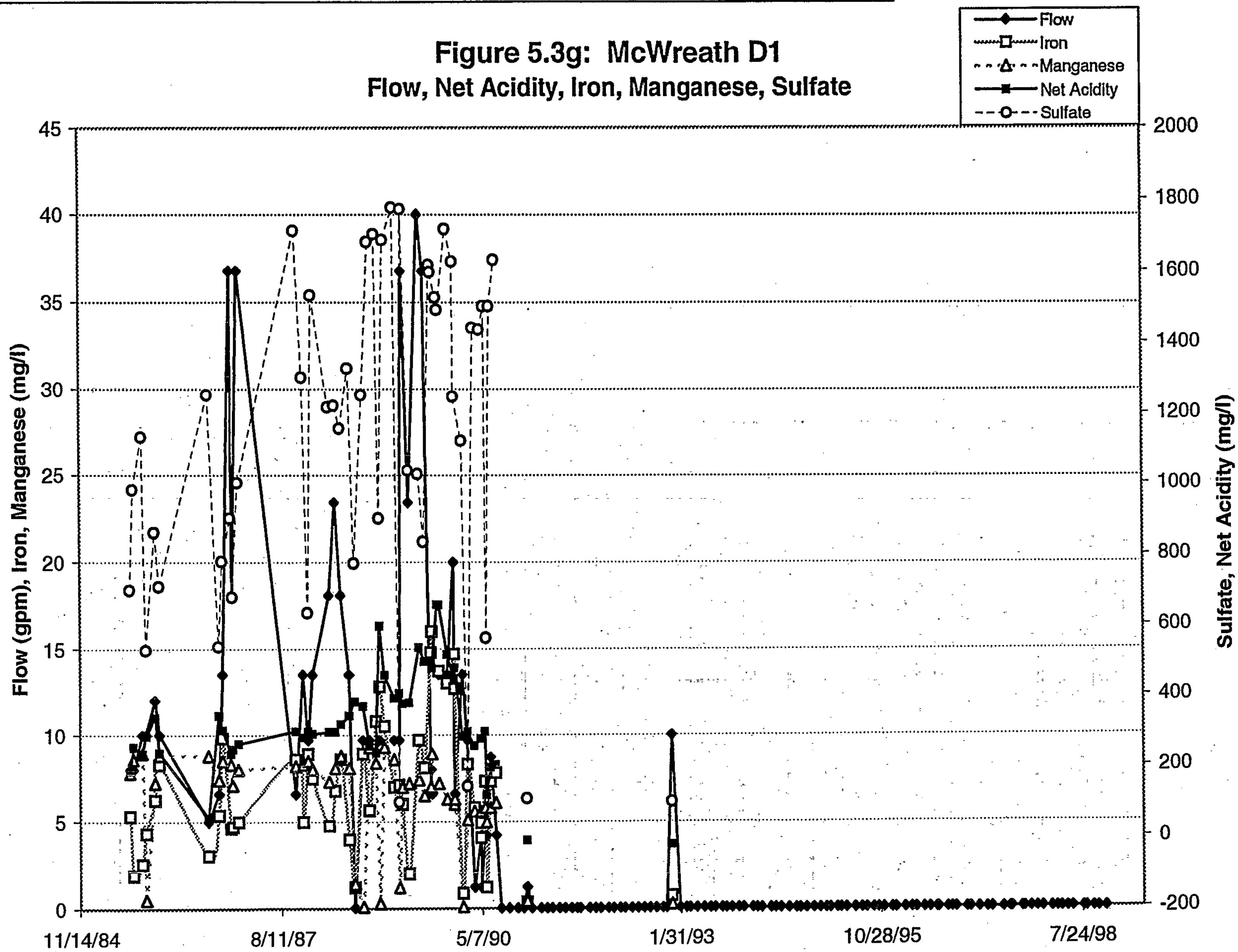


5.3.2 McWreath Remining Site

The McWreath remining site is located in Robinson Township, Washington County, Pennsylvania. The initial surface mining permit for this remining operation was issued on July 21, 1987 for 112.1 acres. The principal best management practice in the remining plan was daylighting of an abandoned underground mine. In this area of Washington County, the overburden of the Pittsburgh Coal includes extensive calcareous strata which produce alkaline mine drainage when disturbed. A similar daylighting example for pH changes at the Solar mine of the Pittsburgh Coal seam, in Allegheny County, Pennsylvania is included in Brady, 1998. The McWreath site had three pre-existing pollution discharges emanating from abandoned underground mine workings prior to remining (monitoring points D-1, D-3, and D-4). The remining operation mined through these discharge locations, and the effects on the flow and water quality are shown in Figures 5.3g through 5.3j.

The flow and concentrations of net acidity, sulfate, iron, manganese, and aluminum in the D-1 discharge are shown in Figure 5.3g. This was the largest of the three deep mine discharges at the McWreath site. This discharge had four sampling events during baseline data collection and following permit issuance when the flow was between 35 and 40 gallons per minute. In April of 1990, the discharge dried up, only briefly reappearing as a 1.2 gallon per minute flow in December 1990, and as a 10 gallon per minute flow in December 1992. According to monthly monitoring data, the discharge has otherwise gone dry as a result of remining from 1990 to present. Flow and concentrations of iron, manganese, acidity, sulfate, and aluminum in the D3 and D4 discharges are shown in Figures 5.3h through 5.3j.

Figure 5.3g: McWreath D1
Flow, Net Acidity, Iron, Manganese, Sulfate



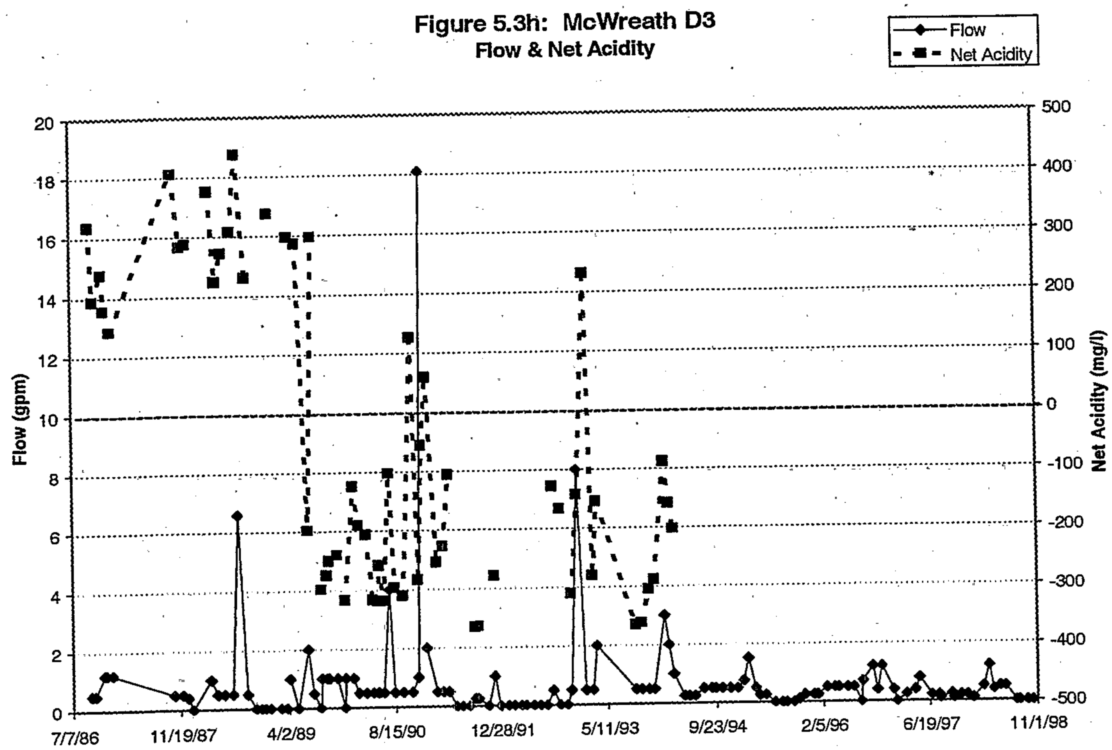


Figure 5.3h: McWreath D3
Flow & Net Acidity
Flow
Net Acidity
Flow (gpm)
Net Acidity (mg/l)
20
18
16
14
12
10
8
6
4
2
0
500
400
300
200
100
0
-100
-200
-300
-400
-500
7/7/86
11/19/87
4/2/89
8/15/90
12/28/91
5/11/93
9/23/94
2/5/96
6/19/97
11/1/98

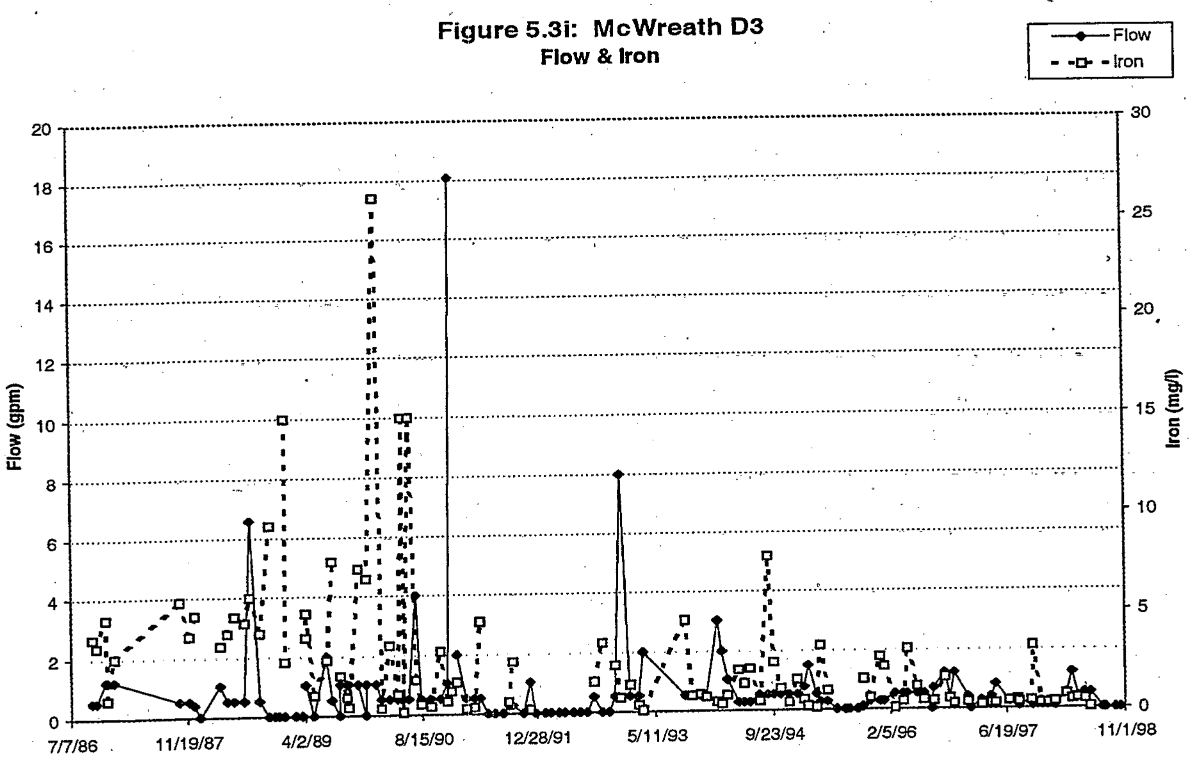
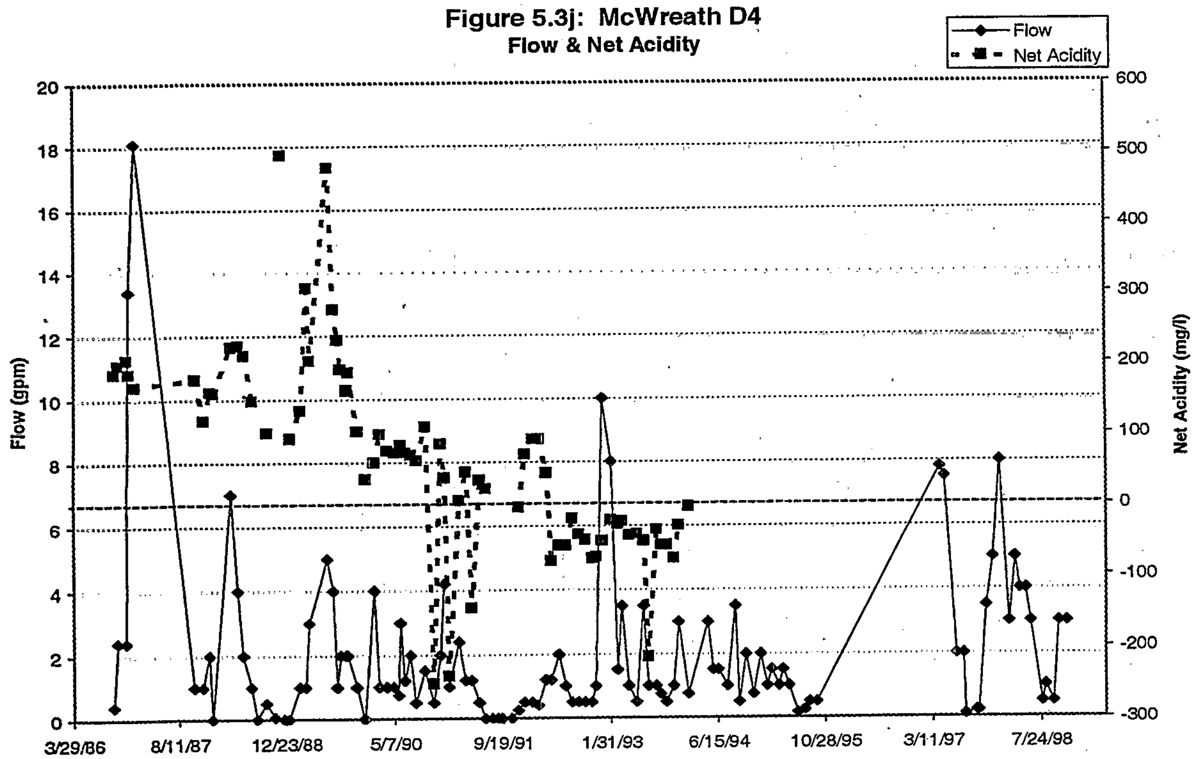


Figure 5.3i: McWreath D3
Flow & Iron
Flow
Iron
Flow (gpm)
Iron (mg/l)
20
18
16
14
12
10
8
6
4
2
0
30
25
20
15
10
5
0
7/7/86
11/19/87
4/2/89
8/15/90
12/28/91
5/11/93
9/23/94
2/5/96
6/19/97
11/1/98

Figure 5.3h shows a dramatic change in net acidity concentration which was in the range of 200 to 400 mg/L in 1986 and 1987, but since 1989, has dropped to predominantly less than 0 and as low as -350 mg/L. Figure 5.3j shows a substantial reduction in iron concentration since 1990. Most of the flow measurements from 1996 through 1999 in Figures 5.3h and 5.3i are less than 2 gallons per minute.

The D-4 discharge had a pre-mining flow intermediate between discharge D-3 and D-1, and recent flow measurements for this discharge are several times higher than the D-3 discharge. Figure 5.3j depicts a large change in net acidity concentration as a result of remining on the McWreath site, where the net acidity prior to 1990 was always greater than 100 mg/L and as high as 500 mg/L, and the net acidity since 1990 is always less than 100 mg/L and as low as -250 mg/L. Therefore, remining transformed two distinctly acidic discharges into distinctly alkaline discharges through daylighting the abandoned deep mine and exposing naturally alkaline overburden strata during remining and reclamation operations.

Figure 5.3j: McWreath D4
Flow & Net Acidity



5.3.3 Trees Mills Site

The Trees Mills remining site is situated in Salem Township, Westmoreland County, Pennsylvania (Figure 5.3k). The remining permit boundary is shown in this figure as a bold line. The surface mining permit for the 325 acre site was issued on May 25, 1990. Surface water drainage from the permit area flows to Beaver Run to the West and Porter Run to the East. Beaver Run is classified as a High Quality – Cold Water Fishery, and the Beaver Run Reservoir (a public water supply impoundment for 100,000 people) is located less than 2500 feet downstream from the Trees Mills remining site.

The primary best management practice in the pollution abatement plan for this site was the daylighting of an abandoned underground mine on the Pittsburgh Coal seam. There were also abandoned surface mine pits and highwalls that were regraded and reclaimed. As the result of extensive mine subsidence overlying the abandoned underground mine, prior to remining, much of the surface of the site resembled a waffle ground that promoted internal drainage to the abandoned deep mine workings rather than overland surface runoff. The geochemical characteristics of the overburden strata were more conducive to acidity production than alkalinity production. Figure 5.3l (Brady et al., 1998) features drill hole data for this site. Geochemical information listed on the left hand side of the bore holes in this figure represent percent sulfur; information listed on the right hand side represent neutralization potential in $CaCO_3$ equivalents. Except for high sulfur shale strata immediately overlying the coal, the overburden strata are characterized by a thick sandstone unit with several zones of relatively high sulfur. Only two sandstone samples in OB-2 have appreciable neutralization potential. The overburden quality of the Pittsburgh Coal at the Trees Mills site is much different (i.e., less calcareous strata, less alkalinity production potential) than at the McWreath site. Hence, the success of the remining pollution abatement plan for the Trees Mills site is more dependent on the hydrogeologic characteristics than on the geochemical characteristics that were significant at the Fisher and McWreath remining sites.

Figure 5.3k: Trees Mills Site Map

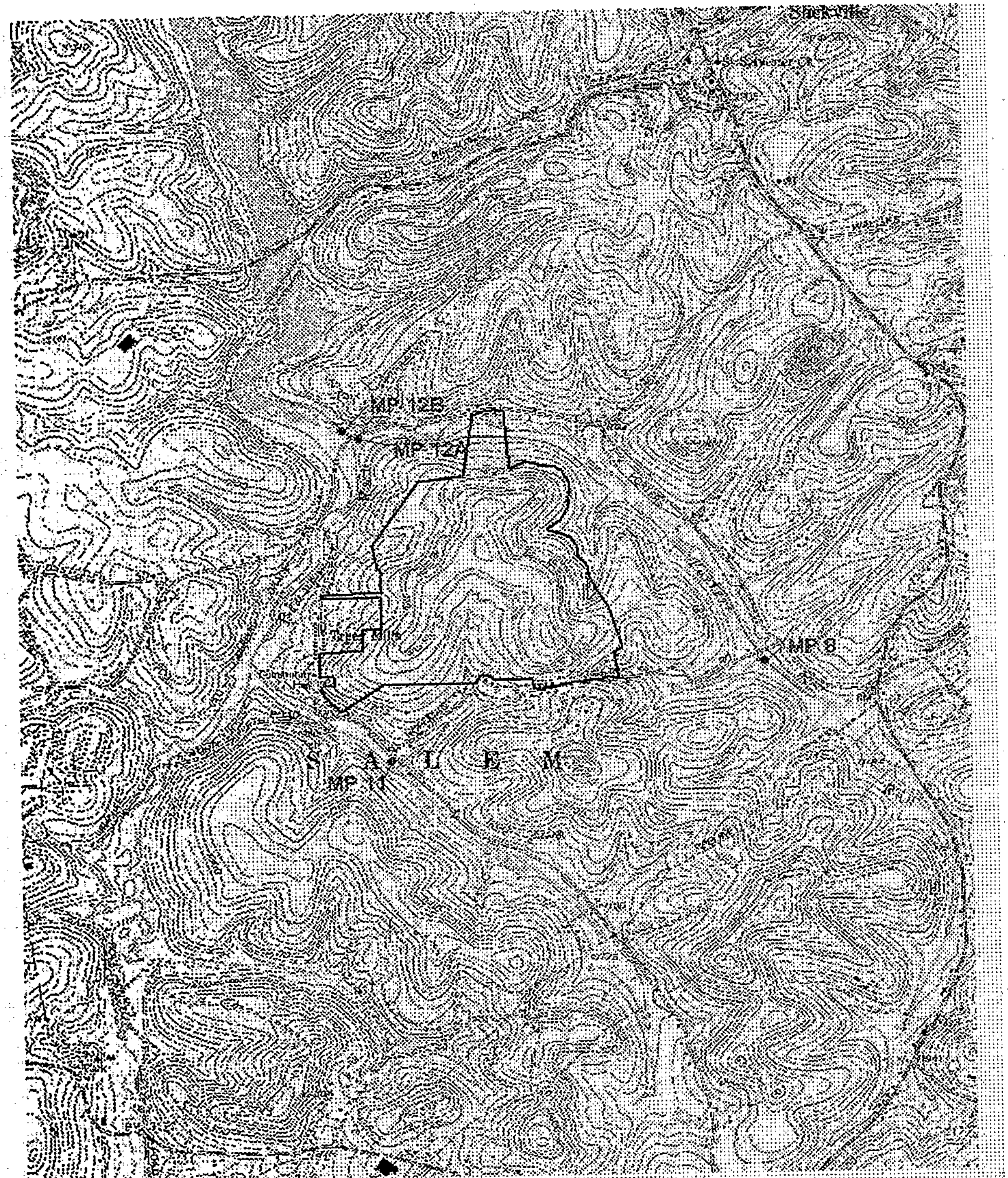
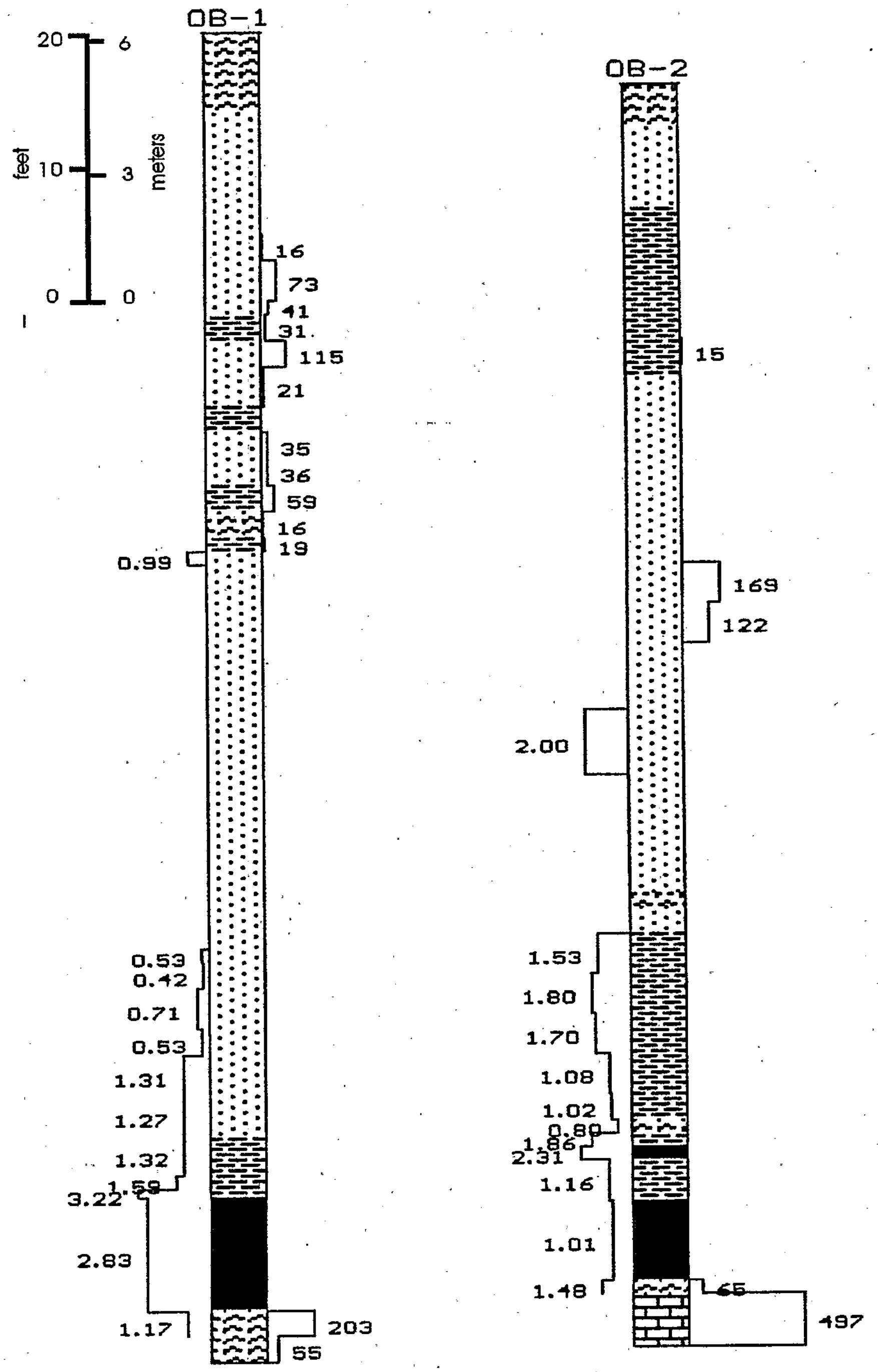


Figure 5.3l: Trees Mills Drill Hole Data

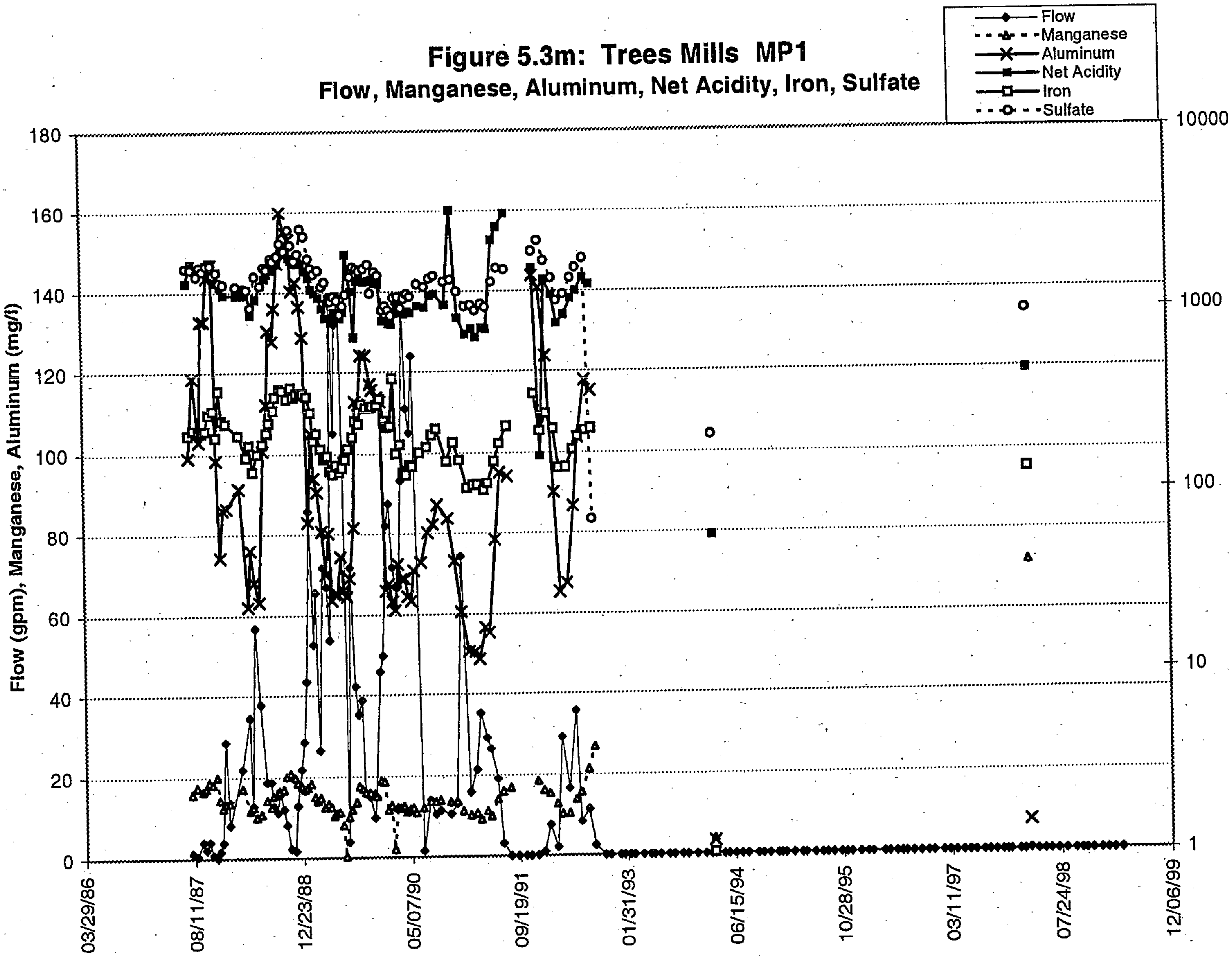


There were numerous acid mine drainage discharges and seeps emanating from the abandoned underground mine workings at the Trees Mills site prior to remining, and baseline pollution load statistical calculations were completed for ten of these monitoring points. The largest of these pre-existing discharges was MP-1 with pre-mining flows as high as 139 gallons per minute (Figure 5.3m and 5.3n). The effects of remining upon three other pre-existing discharges (MP-2, MP-3, and MP-6) will also be discussed. Because, remining operations commenced on the Trees Mills site on October 1991, water quality and flow data from 1987 through September 1991 can be considered pre-mining data. According to mine inspection reports, backfilling was completed by May 14, 1998, thus the intervening time from September 1991 through May 1998 includes the phases of active open pit mining and reclamation activities.

Figure 5.3m shows the variations in flow and concentrations of net acidity, sulfate, iron, manganese, and aluminum in the MP-1 deep mine discharge. The flow was highly variable prior to the initiation of mining in October 1991, ranging from less than one gpm to 139 gpm, with a median flow of 21.7 gpm and an average flow of 38.96 gpm. As the result of remining, the MP-1 discharge dried up by October 1992, reappearing in only one sampling event during the next seven years (0.26 gpm flow on March 3, 1998). The range in acidity concentrations for the period of July 1987 through October 1991 was 773 to 3,616 mg/L with a median of 1,336 mg/L and an average of 1,417 mg/L. The corresponding range in iron concentrations for the MP-1 discharge was 104 to 430 mg/L, with a median of 211 mg/L and an average of 224 mg/L.

Figure 5.3m: Trees Mills MP1

Flow, Manganese, Aluminum, Net Acidity, Iron, Sulfate



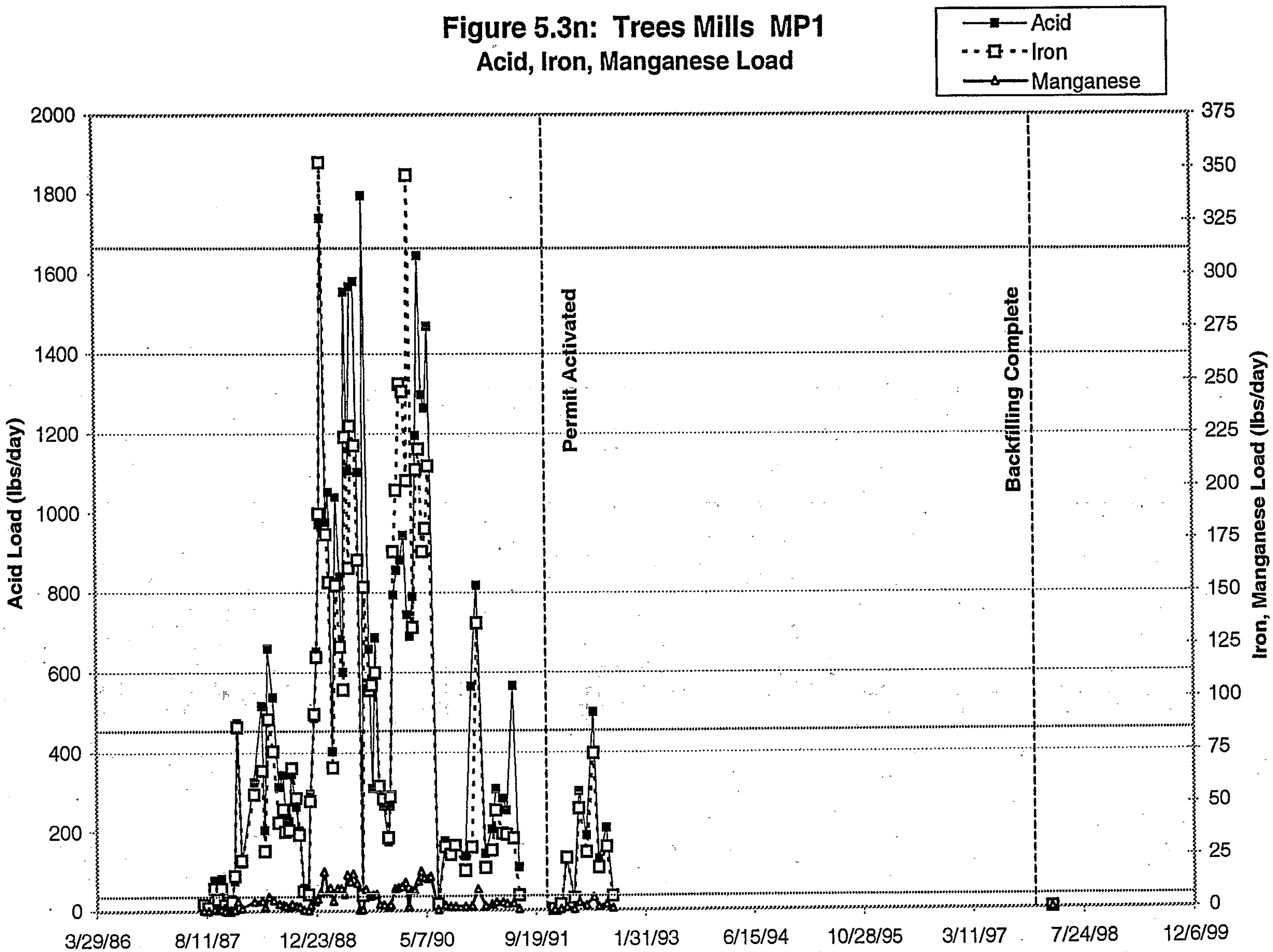


Figure 5.3n: Trees Mills MP1
Acid, Iron, Manganese Load
Acid
Iron
Manganese
Acid Load (lbs/day)
Iron, Manganese Load (lbs/day)
Permit Activated
Backfilling Complete
2000
1800
1600
1400
1200
1000
800
600
400
200
0
375
350
325
300
275
250
225
200
175
150
125
100
75
50
25
0
3/29/86
8/11/87
12/23/88
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
3/11/97
7/24/98
12/6/99

Figure 5.3o: Trees Mills MP2
Flow, Iron, Manganese, Aluminum, Net Acidity, Sulfate

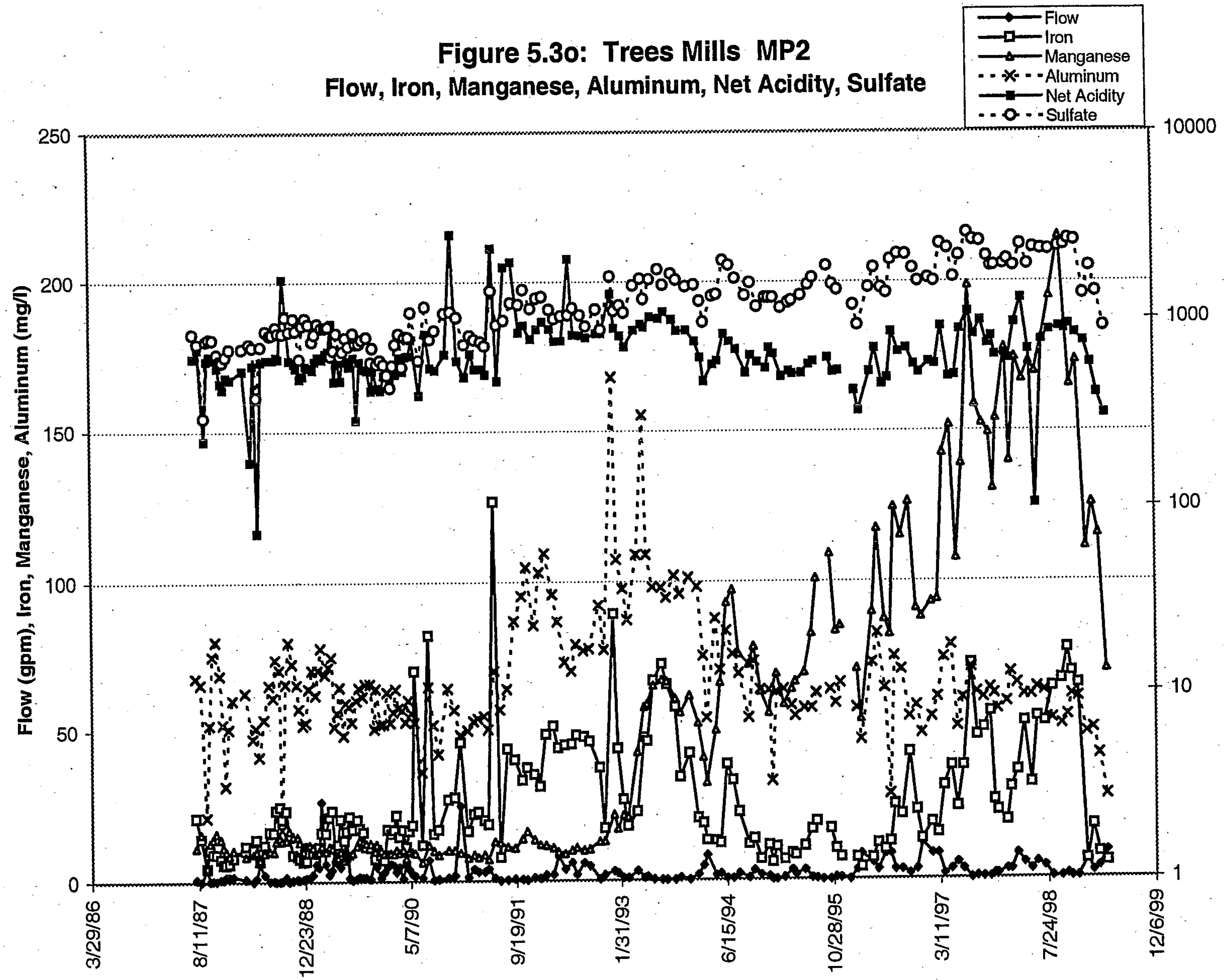


Figure 5.3n shows the variations in the pollution load of acidity, iron, and manganese prior to and following permit activation. The three horizontal dashed lines represent the 95 percent tolerance levels around the frequency distribution of acid loads and the median value of 439 pounds per day acid load calculated in the baseline statistical analysis. The corresponding iron load was 71.8 pounds per day for the baseline sampling period. The median acid load during the period from October 1991 through October 1992 (while the discharge was still flowing) was 128.5 pounds per day and the corresponding iron load was 20.27 pounds per day. The remining operation at the Trees Mills site removed a significant acid load (439 pounds per day = 160,000 pounds per year) and iron load (71.8 pounds per day = 26,000 pounds per year) from the Beaver Run tributary and the Beaver Run public water supply reservoir.

Variations in flow and concentrations of acidity, sulfate, iron, manganese, and aluminum in the MP-2 discharge are shown in Figure 5.3o. Corresponding variations in acidity, iron, and manganese loads prior to permit activation, during mining, and following backfilling are shown in Figure 5.3n. This discharge had substantially lower flow than the MP-1 discharge. The range in flow prior to permit activation was 0.1 to 26.4 gpm (median of 1.3 gpm, average of 3.12 gpm). During active mining, the flow of the MP-2 discharge ranged from 0.02 to 12.1 gpm (median of 1.75, average of 2.66 gpm), while the flow following backfilling ranged from 0.39 to 8.9 gpm (median of 2.55, average of 2.97 gpm). Thus, while the range of flows decreased during mining and post-mining, the median flow increased by approximately one gpm. Net acidity, sulfate, and iron concentrations increased following permit activation (Figure 5.3m). Aluminum concentrations increased during mining but returned to pre-mining levels following backfilling. There also was a notable increase in manganese concentrations, from a median of 10.24 mg/L pre-mining to a median of 171.2 mg/L following backfilling.

The overall environmental impacts of these water quality changes are put in perspective by examining the pollution load data for MP-2 (Figure 5.3n), in comparison to the pollution load reduction achieved at the MP-1 location. The horizontal dashed lines represent the upper and lower 95 percent tolerance levels and the median baseline acid load. Baseline (pre-mining) acid load for MP-2 ranged from 0.31 to 197.6 lbs/day (median of 8.55). Post-backfilling acid load ranged from 3.89 to 54.95 lbs/day. Hence, while extreme values were reduced, median acid load increased by approximately 5 lbs/day. The range in pre-mining iron loads was 0.02 to 13.9 lbs/day (median of 0.26), while the post backfilling range was 0.3 to 3.36 lbs/day (median of 0.46). The range of pre-mining manganese loads was 0.02 to 3.79 lbs/day (median of 0.19), while the post-backfilling range was 0.78 to 10.6 lbs/day (median of 4.31). Based upon median values, there was an increase of 4.31 lbs/day of manganese from the MP-2 discharge, but an elimination of 3.22 lbs/day (average of 5.78 lbs/day) from MP-1. There was a corresponding increase of 0.2 lbs/day iron load from MP-2, with an elimination of 71.8 lbs/day from MP-1. Finally there was an increase of approximately 4.6 lbs/day acid load from MP-2, offset by the elimination of 439 lbs/day from MP-1.

The net effect on the Beaver Run receiving stream was a significant reduction in pollution loads. Variations in concentrations of net acidity, sulfate, iron, manganese, and aluminum from MP-3 (Figure 5.3q) are similar to that from MP-2, except for a significant reduction in iron concentration. Pre-mining iron concentrations in MP-3 ranged from 7.9 to 226.4 mg/L (median of 75), while the post-backfilling iron concentrations ranged from 6.55 to 84.72 mg/L (median value of 29.77). Pre-mining median manganese concentration was 11.18 mg/L, and the post-backfilling median was 194.55. Aluminum concentrations were 61.93 mg/L pre-mining and 63.38 mg/L post-backfilling. The flow of MP-3 ranged from 0.1 to 67 gpm pre-mining (median of 6.95), while post-backfilling flow ranged from 1.5 to 21.7 gpm (median of 3.95). Variations in acidity, iron, and manganese loads from MP-3 are shown in Figure 5.3r. Again, the horizontal dashed lines represent the upper and lower 95 percent tolerance levels and median value for pre-mining acid loads. The median baseline pollution load for acidity is 41.79 lbs/day compared to a post-backfilling median acid load of 34.79 lbs/day. The corresponding medians for iron loads are

Figure 5.3p: Trees Mills MP2
Acid, Iron, Manganese Load

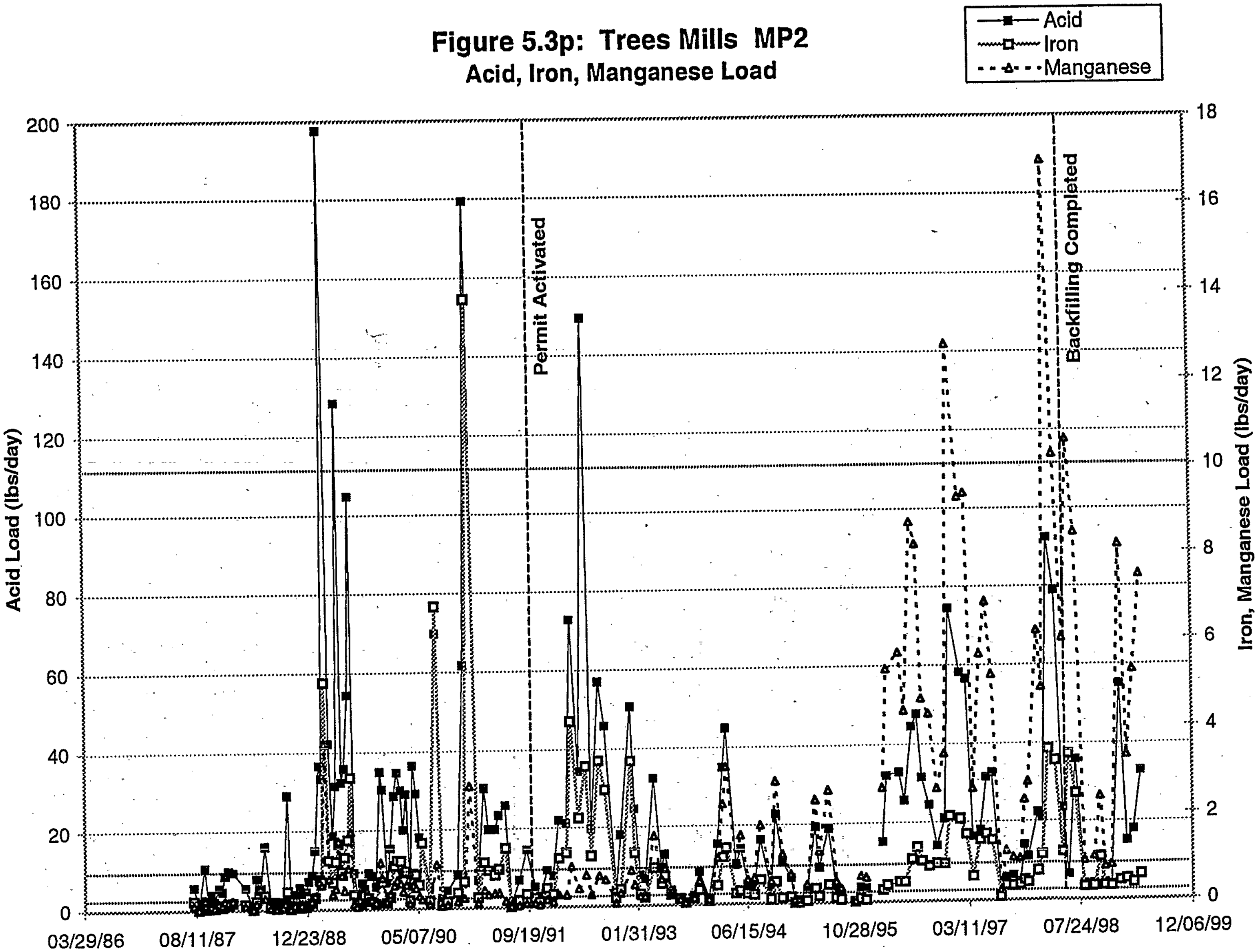
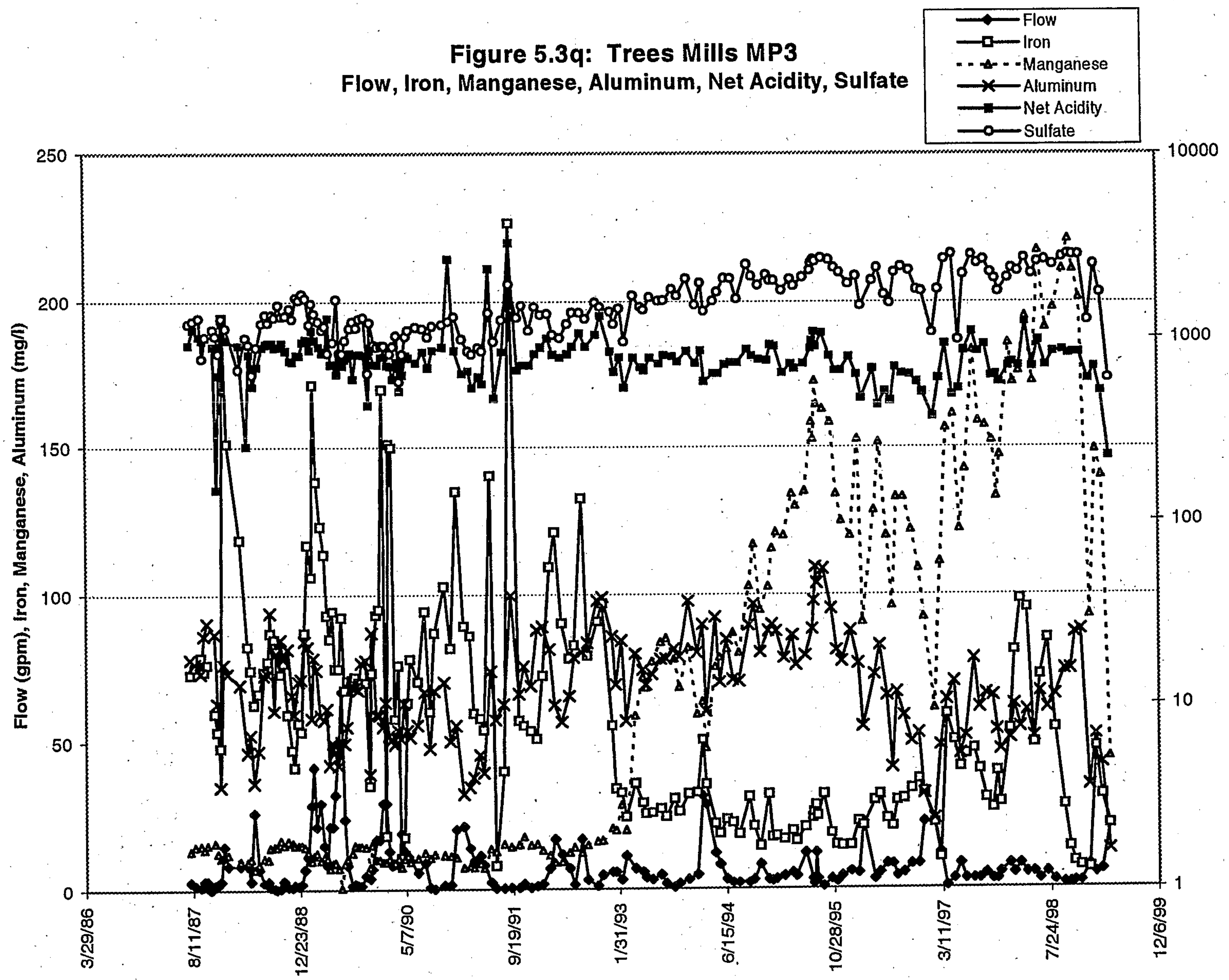


Figure 5.3q: Trees Mills MP3
Flow, Iron, Manganese, Aluminum, Net Acidity, Sulfate



4.03 lbs/day pre-mining and 1.95 following backfilling. The pre-mining median manganese load was 0.7 lbs/day, and increased to a median of 7.97 lbs/day post-backfilling. Extreme values of iron loads were substantially reduced following backfilling.

The MP-6 discharge is located below the outcrop of the Pittsburgh Coal seam, and varied in pre-mining flow from 0.8 to 62.4 gpm (median of 9.5). Following completion of backfilling, the range in flow is 0.39 to 21.7 gpm (median of 3.8). The pre-mining range of acidity concentration from the MP-6 discharge was 125 to 2,587 mg/L (median of 784.7), and the post-mining range was 522 to 968 mg/L (median of 804.5). The range of the iron concentrations pre-mining was 11.54 to 161.0 mg/L (median of 74.7), while the post-mining range was 31.64 to 94.73 mg/L (median of 55.62). The pre-mining range in manganese concentration was 6.62 to 19.24 mg/L (median of 11.3), while the post-mining manganese range was 14.63 to 30.14 mg/L (median of 26.06 mg/L). Again, the horizontal dashed lines in Figure 5.3u represent the median acid load and upper and lower 95 percent tolerance levels for baseline statistical calculations. The median acid load was 73.13 lbs/day pre-mining as compared to 38.08 lbs/day post-mining. The corresponding iron loads were a median of 7.5 pre-mining and 2.98 lbs/day post-mining. The pre-mining median load of manganese was 1.11 lbs/day, and was nearly equal to the post-mining median of 1.06 lbs/day.

Due to the cumulative effects of remining upon the MP-1, MP-2, MP-3, and MP-6 discharges, the Trees Mills remining operation has resulted in a significant reduction in the pollution load of acidity and metals (iron, manganese, and aluminum) to the receiving stream and the Beaver Run Reservoir. To determine whether these pollution reduction effects could be detected in the water chemistry of the receiving stream, the permittee's self monitoring reports and PA DEP mining inspector's monitoring data were evaluated from the same monitoring points located upstream and downstream of the Trees Mills operation on the Porter Run and Beaver Run tributaries. The downstream monitoring points are located immediately above the confluence of these two tributaries (MP-12a and MP-12b). The upstream monitoring points are shown in Figure 5.3k.

Figure 5.3r: Trees Mills MP3
Acid, Iron, Manganese Load

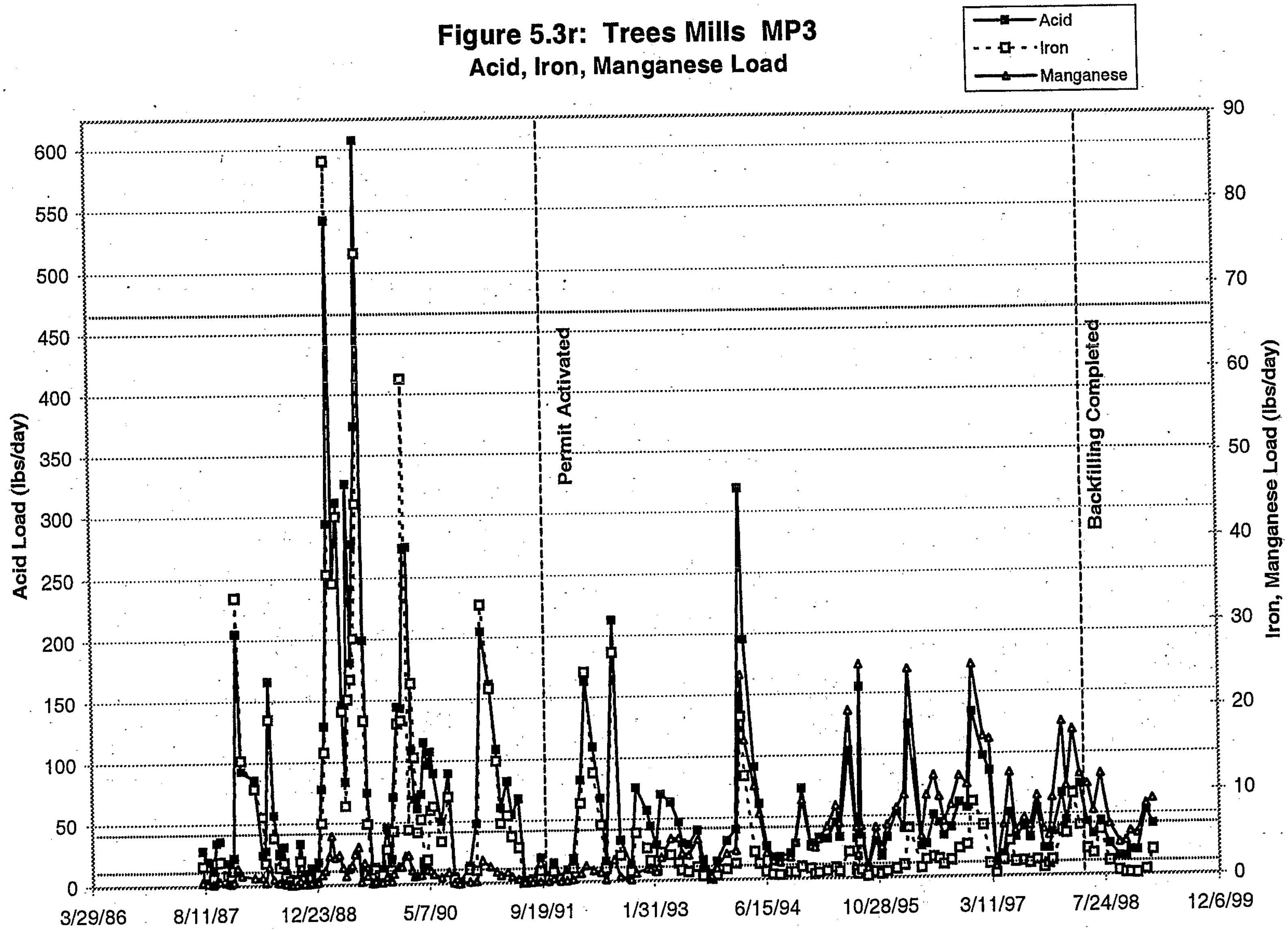
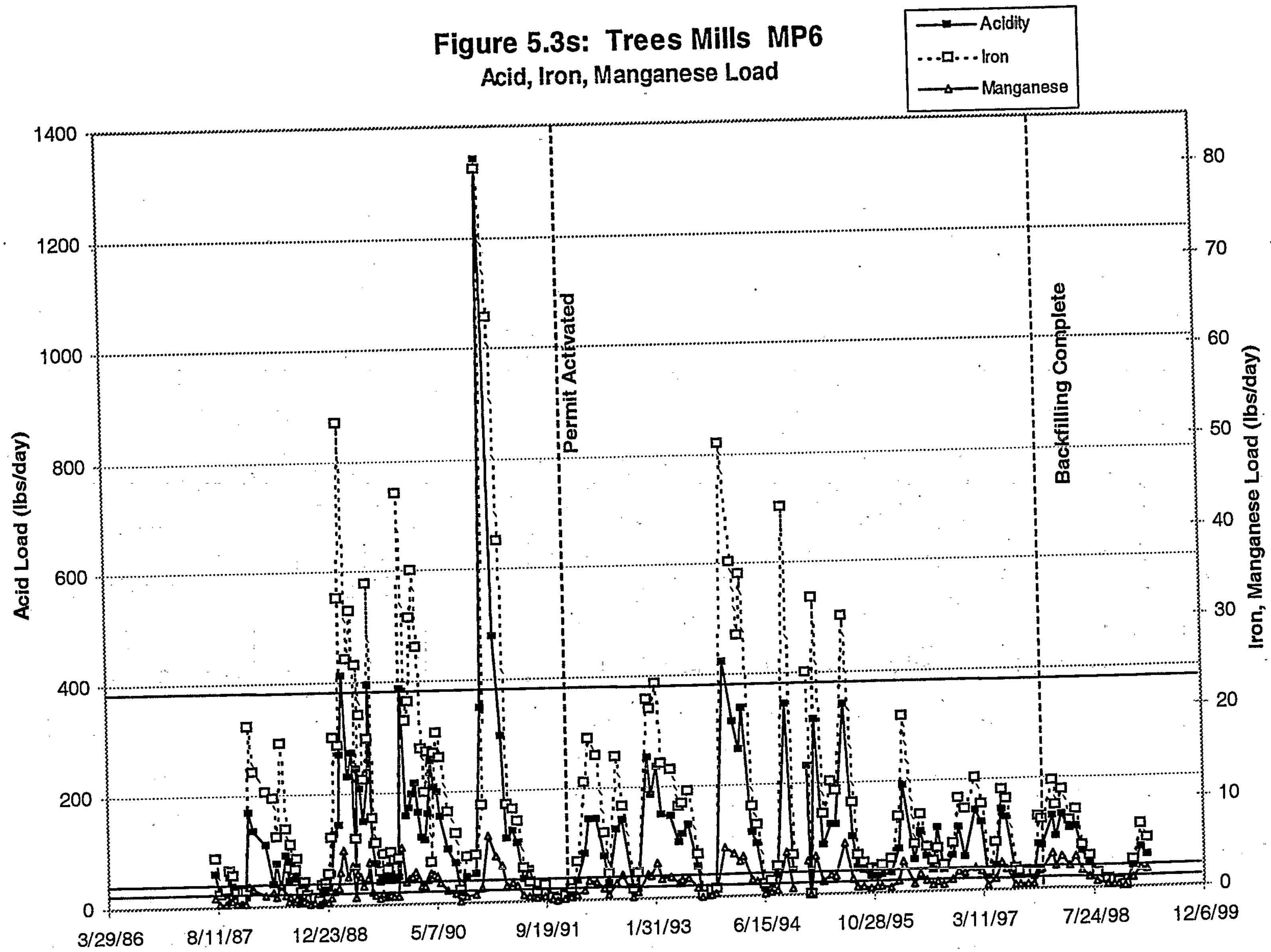
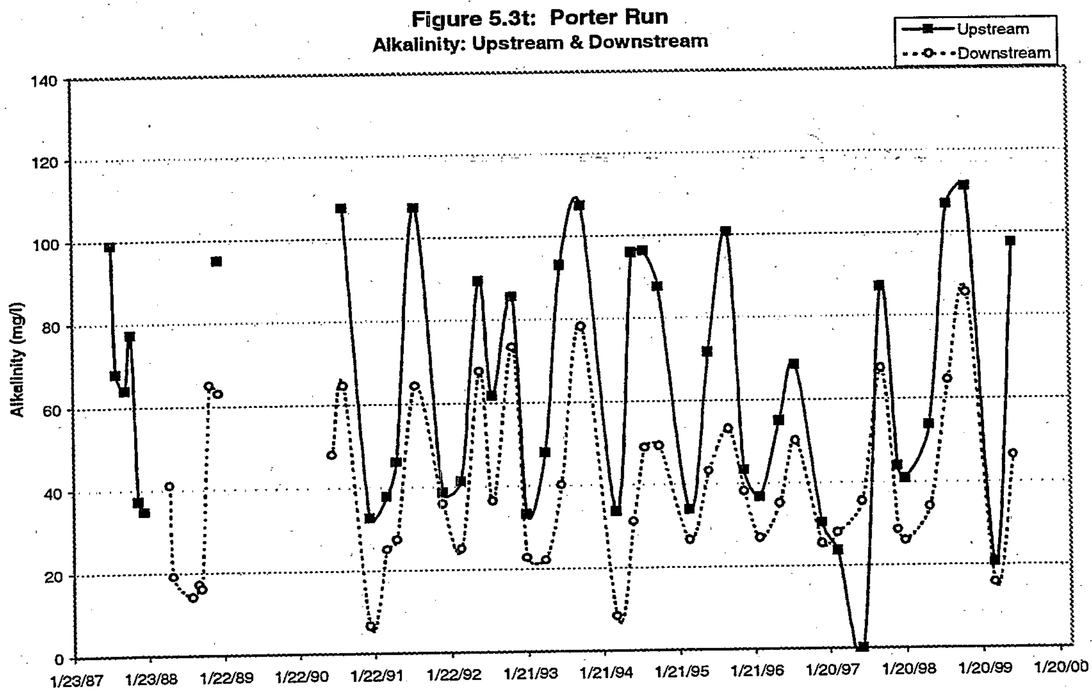


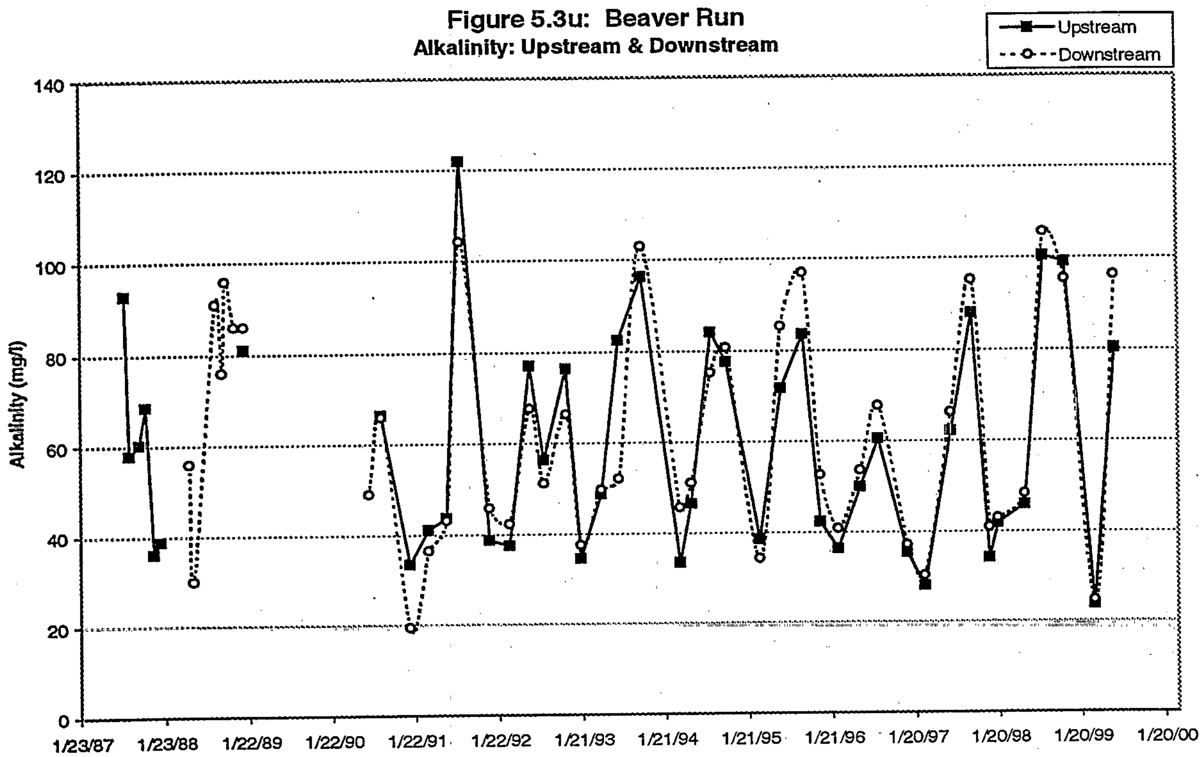
Figure 5.3s: Trees Mills MP6
Acid, Iron, Manganese Load



These two tributaries had appreciable alkalinity concentrations during the entire monitoring period (1987 through 1999), undoubtedly due to the presence of significant carbonate lithologic units in the drainage basin (e.g., the limestone underlying the Pittsburgh Coal Seam, Figure 5.31). However, by comparing upstream and downstream alkalinity concentrations in Porter's Run and Beaver Run (Figures 5.3t and 5.3u), the subtle changes in alkalinity concentration observed are believed to be due to the reduction in acid load from the Trees Mills remining operation. In Figure 5.3t, the upstream alkalinity concentration in Porter's Run is consistently higher than the downstream alkalinity concentration during the period of record. In Beaver Run (Figure 5.3u), the upstream alkalinity concentration was higher than the downstream alkalinity pre-mining and during the first year or two of remining. However, since 1994 the trend reversed, and the downstream alkalinity concentrations are typically higher than the upstream alkalinity concentrations. It is inferred from this data that the MP-1 discharge (and other pollutional discharges) impacted the receiving stream, but the elimination or reduction of pollution loads from these discharges during and following remining increased the downstream alkalinity. The effect of this elimination, likely would be more dramatic without the presence of significant in-stream alkalinity and flow.

Figure 5.3t: Porter Run
Alkalinity: Upstream & Downstream





5.4 Conclusions

- Pre-existing discharges vary widely in flow and consequently, also in pollutant loading rates. Because there is such a large seasonal component to flow variability, it is necessary that baseline pollution load monitoring cover the entire range of seasonal conditions (generally an entire water year). Use of a partial water year may significantly under or over represent the baseline pollution load and therefore is not recommended.

- Not all discharges behave in a similar fashion. Some discharges respond steadily, with relatively small variation, while others change rapidly and by several orders of magnitude. While it is important to consider these behaviors, possibly requiring case-by-case monitoring, most discharges exhibit fairly predictable behavior, and are appropriately monitored using a monthly sampling interval and a one-year baseline monitoring period.

- Although it maybe possible to miss the most extreme high-flow events using a monthly sampling interval, as long as a consistent sample interval is used for determining and monitoring baseline, and the statistical test is not overly sensitive to extreme values, this sampling protocol should be adequate. Low flow events occur over longer periods of longer duration and should be adequately represented with monthly sampling.

- Extremely dry or extremely wet years may pose difficulties in establishing a representative baseline pollution load, but significant year-to-year variations in pollution load are rare and would be even more rare for multiple consecutive years. Seldom would it be worth the additional time and expense to require a multi-year baseline period. However, water quality monitoring should consider the possibility, though infrequent, of year-to-year pollution load variations that rise to the level of statistical significance.

- Remining-induced changes in pollution load tend to be very dramatic and can result from either significant changes in flow or significant changes in water quality. The fact that these changes are rarely subtle makes it relatively easy to design a monitoring program that can detect significant changes, while minimizing the incidence of false positives (i.e., indications of significant changes in water quality which may be due to seasonal changes or changes due to weather patterns). The monthly monitoring interval used in the case studies did adequately document pre and post-remining water quality, and was sufficient to detect significant changes in pollution loading rates.

- Less frequent water monitoring intervals are much more likely to over or under represent the baseline pollution load, and to inaccurately detect changes in pollution loading rates. Monitoring intervals that are more frequent than monthly, are generally unnecessary and may not be worth the added expense.

- Acidity and alkalinity, pH, metals, sulfates, and flow rates, often respond differently depending on the BMP used. Some BMPs may reduce flows while leaving pollutant concentrations unchanged. Sources of alkalinity may increase pH and reduce acidity, increase one or more metals and decrease others, and increase or decrease sulfates. Observing the response of individual parameters allows the analysis of BMP efficiency. This is useful for applying particular BMPs to similar situations, in troubleshooting, and in adding or modifying BMPs to achieve a desired result.

References

Ballaron, P.B., 1999. Water Balance for the Jeddo Tunnel Basin, Luzerne County, Pennsylvania. Publication No. 208. Pennsylvania Department of Environmental Protection, Pottsville District Mining Office under Grant ME97105.

Ballaron, P.B., C.M. Kocher, and J.R. Hollowell, 1999. Assessment of Conditions Contributing Acid Mine Drainage to the Little Nescopeck Creek Watershed, Luzerne County, Pennsylvania, and Abatement Plan to Mitigate Impaired Water Quality in the Watershed. Publication No. 204. Pennsylvania Department of Environmental Protection, Bureau of Watershed Conservation under Grant ME96114.

Barnes, I., W.T. Stuart, and D.W. Fisher, 1964. Field Investigation of Mine Waters in the Northern Anthracite Field, Pennsylvania. U.S. Geological Survey, Professional Paper 473-B. U.S. Government Printing Office: Washington, DC.

Brady, K.T., R.J. Hornberger, and G. Fleeger, 1998. Influence of Geology on Postmining Water Quality: Northern Appalachian Basin. Chapter 8 in Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Edited by K.B.C. Brady, M.W. Smith and J. Schueck. Pennsylvania Department of Environmental Protection: Harrisburg, PA. pp. 8-1 to 8-92.

Cravotta, C.A., 1998. Oxic Limestone Drains for Treatment of Dilute, Acidic Mine Drainage. Paper presented at the 1998 Annual Symposium of the West Virginia Surface Mine Drainage Task Force, Morgantown, WV, April 7-8.

Cravotta, C.A. and M.K. Trahan, 1999. Limestone Drains to Increase pH and Remove Dissolved Metals from Acidic Mine Drainage. Applied Geochemistry, Vol.14, pp. 581-601.

Cravotta, C.A., 1999. Personal communication between Charles Cravotta and Roger Hornberger of the Pennsylvania Department of Environmental Protection, Pottsville, PA.

Fishel, D.K., and J.E. Richardson, 1986. Results of a Preimpoundment Water-Quality Study of

Swatara Creek Pennsylvania. Water Resources Investigations Report 85-4023. US Geological Survey: Harrisburg, PA.

Griffiths, J.C., (No date). Report No. 1: Reconnaissance into the use of MINITAB using Hamilton 01, 08 Data. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987a. Report No. 2: Time Series Analysis of Data from Hamilton 08. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987b. Report No. 3: Analysis of Data from Arnot 001, 003, 004. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987c. Report No. 4: Analysis of Data from the Clarion Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987d. Report No. 5: Analysis of Data from Ernest Refuse Pile, Indiana County, Pennsylvania. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1987e. Report No. 6: Analysis of Data from the Fisher Deep Mine. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988a. Report No. 7: Analysis of Data from the Markson Site. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Griffiths, J.C., 1988b. Report No. 8: Synopsis of Reports 1 - 7. U.S. Environmental Protection Agency, Office of Water: Washington, DC.

Growitz, D.J., et al. 1985. Reconnaissance of Mine Drainage in the Coal Fields of Eastern Pennsylvania. Water Resources Investigations Report 83-4274. US Geological Survey: Harrisburg, PA.

Hornberger, R.J., M.W. Smith, A.E. Friedrich, and H.L. Howell, 1990. Acid Mine Drainage from Active and Abandoned Coal Mines in Pennsylvania. Chapter 32 in Water Resources in Pennsylvania: Availability, Quality and Management. Edited by S.K. Majumdar, E.W. Miller and R.R. Parizek, Easton: The Pennsylvania Academy of Science, pp. 432 - 451.

McCarren, E.F., J.W. Wark, and J.R. George, 1961. Hydrologic Processes Diluting and Neutralizing Acid Streams of the Swatara Creek Basin, Pennsylvania. Geological Survey Research, pp. D64-D67.

Smith, M.W., 1988. Establishing Baseline Pollutant Load from Pre-existing Pollutional

Discharges for Remining in Pennsylvania, Mine Drainage and Surface Mine Reclamation. Paper presented at the 1988 Mine Drainage and Surface Mine Reclamation Conference, Pittsburgh, PA, April 17-22, pp. 311-318.

U.S. EPA, 2001. Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits. Prepared for US Environmental Protection Agency, Office of Water by Pennsylvania Department of Environmental Protection and DynCorp I&ET. EPA-821-B-01-014.

Appendix A: Example Calculations of Statistical Methods

The following calculations are examples of the statistical procedures presented in Section 3.

- Example 1 includes concentration and flow results for 12 baseline and monitoring samples. This example presents application of both Method 1 and Method 2, using the steps defined for 12 samples.
- Example 2 includes loading results for 18 baseline and monitoring samples. This example presents application of both Method 1 and Method 2, and includes the calculation of the Single Observation Trigger specified for 17 samples or greater in Method 1.
- Example 3 includes concentration and flow results for 12 baseline and monitoring samples. This example presents application of both Method 1 and Method 2 and demonstrates a recommended approach when replacing baseline concentrations below the limits established in 40 CFR part 434, Subpart C.

1.0 Example 1

Assume 12 baseline flow and iron concentrations are collected by sampling once per month for a year. Likewise, 12 flow and iron monitoring observations are obtained by sampling once per month for a period of one year. Determination of exceedances are presented using both Methods 1 and 2. For all calculations in Example 1, assume the following flows (in gpm) and iron concentrations (in mg/L).

Flow

Baseline	5.0	14.0	42.0	35.0	26.0	22.0	12.0	11.0	11.0	6.0	11.0	6.0
Monitoring	8.0	14.0	15.0	32.0	43.0	28.0	16.0	14.0	16.0	9.0	9.0	11.0

Iron Concentration

Baseline	14.2	14.0	20.6	13.6	13.2	12.4	13.2	13.4	13.6	14.3	15.2	13.4
Monitoring	16.2	17.8	16.3	15.1	20.9	15.7	15.8	16.7	15.4	15.6	16.7	15.3

The resulting iron loads (in lbs/day = flow * concentration * 0.01202) are given below:

Iron Loads

Baseline	0.85	2.36	10.40	5.72	4.13	3.28	1.90	1.77	1.80	1.03	2.01	0.97
Monitoring	1.56	3.00	2.94	5.81	10.80	5.28	3.04	2.81	2.96	1.69	1.81	2.02

1.1 Single Observation Trigger:

1.1.1 Method 1 (See Figure 3.2a):

1) Twelve baseline observations were collected, therefore n = 12.

2) The baseline loading observations are placed in sequential order from smallest to largest. [0.85, 0.97, 1.03, 1.77, 1.80, 1.90, 2.01, 2.36, 3.28, 4.13, 5.72, 10.40]

3) The number of observations, n, is less than 16, therefore the Single Observation Trigger (L) equals $x_{(12)}$ (the maximum) = 10.40.

4) One monitoring load (10.80) is greater than 10.40, therefore the Single Observation Trigger (L) was exceeded.

1.1.2 Method 2 (See Figure 3.2b):

1) Twelve is an even number, therefore the median of the modified baseline observations is:
   $M = 0.5 * (x_{(6)} + x_{(7)})$.
   $M = 0.5 * (1.90 + 2.01) = 1.955$

   In order to determine M_1, calculate the median of the subset ranging from $x_{(7)}$ to $x_{(12)}$. Because 12 - 6 = 6 is even, $M_1 = 0.5 * (x_{(9)} + x_{(10)})$
   $M_1 = 0.5 * (3.28 + 4.13) = 3.705$

2) In order to determine M_{-1}, calculate the median of the subset ranging from $x_{(1)}$ to $x_{(6)}$. Because 6 is even, $M_{-1} = 0.5 * (x_{(3)} + x_{(4)})$
   $M_{-1} = 0.5 * (1.03 + 1.77) = 1.40$

3) To calculate R, subtract M_{-1} from M_1.
   R = 3.705 - 1.40 = 2.305

4) $L = M_1 + (3 * R) = 3.705 + (3 * 2.305) = 10.62$

5) One monitoring observation (10.80) is greater than 10.62, therefore the Single Observation Trigger (L) was exceeded.

1.2 Annual Comparison

1.2.1 Method 1 (See Figure 3.2a)

1) From 1.1.2 Step 1, $M_1 = 3.705$

2) From 1.1.2 Step 2, $M_{-1} = 1.40$

3) From 1.1.2 Step 3, R = 2.305

4) The calculated value for R is then substituted into the equation for T.

$$T = 1.955 + \frac{1.815 * 2.305}{\sqrt{12}} = 3.16$$

5) The following monitoring observations are ordered from smallest to largest.
[1.56, 1.69, 1.81, 2.02, 2.81, 2.94, 2.96, 3.00, 3.04, 5.28, 5.81, 10.80]

6) There are 12 monitoring observations, therefore m = 12.
The number of observations is even, therefore $M' = 0.5 * (x_{(6)} + x_{(7)})$
$M' = 0.5 * (2.94 + 2.96) = 2.95$
This holds true for M_1' and M_{-1}' as well.
$M_1' = 0.5 * (x_{(9)} + x_{(10)}) = 0.5 * (3.04 + 5.28) = 4.16$
$M_{-1}' = 0.5 * (x_{(3)} + x_{(4)}) = 0.5 * (1.81 + 2.02) = 1.915$

7) To calculate R, subtract M_{-1}' from M_1'
$R' = 4.16 - 1.915 = 2.245$.

8) The calculated value for R' is then substituted in the equation for T '.

$$T' = 2.95 - \frac{1.815 * 2.245}{\sqrt{12}} = 1.77$$

9) T' (1.77) is less than T (3.16), therefore the median baseline pollution loading was not exceeded.

1.2.2 Method 2 (Wilcoxon-Mann-Whitney Test) (See Figure 3.2b)

Instructions for the Wilcoxon-Mann-Whitney test are given in Conover (1980), cited in Figure 3.2b.

1) When using both baseline and monitoring data, n = 12 and m =12

2) The baseline and monitoring observations are listed with their corresponding rankings.

Baseline Observations (lbs/day)	0.85	0.97	1.03	1.77	1.80	1.90	2.01	2.36	3.28	4.13	5.72	10.40
Baseline Rankings	1	2	3	6	7	9	10	12	18	19	21	23
Monitoring Observations (lbs/day)	1.56	1.69	1.81	2.02	2.81	2.94	2.96	3.00	3.04	5.28	5.81	10.80
Monitoring Rankings	4	5	8	11	13	14	15	16	17	20	22	24

3) The sum of the twelve baseline ranks (S_n) = 131.

4) In order to find the appropriate critical value (C), match the column with the correct n (number of baseline observations) to the row with the correct m (number of monitoring observations). As found in the table, the critical value C for 12 baseline and 12 monitoring observations is 99.

Critical Values (C) of the Wilcoxon-Mann-Whitney Test
(for a one-sided test at the 99.9 percent level)

n / m	10	11	12	13	14	15	16	17	18	19	20
10	66	79	93	109	125	142	160	179	199	220	243
11	68	82	96	112	128	145	164	183	204	225	248
12	70	84	99	115	131	149	168	188	209	231	253
13	73	87	102	118	135	153	172	192	214	236	259
14	75	89	104	121	138	157	176	197	218	241	265
15	77	91	107	124	142	161	180	201	223	246	270
16	79	94	110	127	145	164	185	206	228	251	276
17	81	96	113	130	149	168	189	211	233	257	281
18	83	99	116	134	152	172	193	215	238	262	287
19	85	101	119	137	156	176	197	220	243	268	293
20	88	104	121	140	160	180	202	224	248	273	299

5) S_n (131) is greater than C (99). Therefore, according to the Wilcoxon-Mann-Whitney Test, the monitoring observations did not exceed the baseline pollution loading.

2.0 Example 2

Assume 18 baseline iron loading determination observations are collected by sampling twice per month for nine months. Likewise, 18 iron load monitoring observations are obtained by sampling twice per month for a period of nine months. Examples of both Methods 1 and 2 are presented below. For all calculations in Example 2, assume the following iron load observations (in lbs/day):

Observation	Baseline	Monitoring
1	0.030	0.530
2	0.005	0.040
3	1.915	1.040
4	0.673	0.033
5	0.064	0.030
6	0.063	0.230
7	0.607	0.710
8	0.553	0.240
9	0.286	0.390
10	0.106	0.830
11	0.406	3.050
12	1.447	0.580
13	0.900	1.180
14	0.040	0.510
15	2.770	0.046
16	1.803	0.690
17	0.160	0.630
18	0.045	0.370

<u>2.1 Single Observation Trigger</u>

2.1.1 Method 1 (See Figure 3.2a)

1) The number of baseline observations collected, n = 18.

2) The baseline observations are ordered sequentially from smallest to largest. [0.005, 0.030, 0.040, 0.045, 0.063, 0.064, 0.106, 0.160, 0.286, 0.406, 0.553, 0.607, 0.673, 0.900, 1.447, 1.803, 1.915, 2.770]

3) The number of observations is greater than 16, therefore M, M_1, M_2 and M_3 must be calculated. The number of observations is even, which means the median of the baseline observations must be calculated using the following equation:

$M = 0.5 * (x_{(9)} + x_{(10)})$.
$M = 0.5 * (0.286 + 0.406) = 0.346$

4) To determine M_1, calculate the median of the subset ranging from $x_{(10)}$ to $x_{(18)}$.
$18 - 9 = 9$ is odd, therefore $M_1 = x_{(14)} = 0.900$.

5) To determine M_2, calculate the median of the subset ranging from $x_{(14)}$ to $x_{(18)}$.
$18 - 13 = 5$ is odd, therefore $M_2 = x_{(16)} = 1.803$.

6) To determine M_3, calculate the median of the subset ranging from $x_{(16)}$ to $x_{(18)}$.
$18 - 15 = 3$, which is odd, therefore $M_3 = x_{(17)} = 1.915$.

7) To determine L, calculate the median of the subset ranging from $x_{(17)}$ to $x_{(18)}$.
$18 - 16 = 2$, which is even, therefore $L = 0.5 * (x_{(17)} + x_{(18)}) = 0.5 * (1.915 + 2.770) = 2.343$.

8) One monitoring observation (3.050) is above L (2.343), therefore the Single Observation Trigger was exceeded.

2.1.2 Method 2 (See Figure 3.2b)

1) From 2.1.1 Step 4, M_1 (the third quartile of the baseline data) is equal to 0.900.

2) To find M_{-1}, calculate the median of the subset ranging from $x_{(1)}$ to $x_{(9)}$.
9 is odd, therefore $M_{-1} = x_{(5)} = 0.063$.

3) The value for R is found by subtracting M_{-1} from M_1
$R = 0.900 - 0.063 = 0.837$

4) $L = M_1 + (3 * R) = 0.900 + (3 * 0.837) = 3.411$.

5) All monitoring observations are less than 3.411, therefore the Single Observation Trigger (L) was not exceeded.

2.2 Annual Comparison

2.2.1 Method 1 (See Figure 3.2a)

1) As determined in Section 2.1.2 step 1, M = 0.346, and $M_1 = 0.900$.

2) As determined in Section 2.1.2 step 2, $M_{-1} = 0.063$.

3) As determined in Section 2.1.2 step 3, R = 0.837.

4) To find T, the value for R is inserted in the following equation:

$$T = 0.346 + \frac{1.815 * 0.837}{\sqrt{18}} = 0.704$$

5) The monitoring observations are placed in order from lowest to highest.
[0.030, 0.033, 0.040, 0.046, 0.230, 0.240, 0.370, 0.390, 0.510, 0.530, 0.580, 0.630, 0.690, 0.710, 0.830, 1.040, 1.180, 3.050]

6) The number of monitoring observations (m) = 18.
18 is even, making $M' = 0.5 * (x_{(9)} + x_{(10)})$
$M' = 0.5 * (0.510 + 0.530) = 0.520$

7) To determine M_1', calculate the median of subset $x_{(10)}$ to $x_{(18)}$.
Because 18 - 9 = 9 is odd, $M_1' = (x_{(14)}) = 0.710$

8) To determine M_{-1}', calculate the median of subset $x_{(1)}$ to $x_{(9)}$.
Because 9 is odd, $M_{-1}' = (x_{(5)}) = 0.230$

9) The value for R' is found by subtracting M_{-1}' from M_1'.
R' = 0.710 - 0.230 = 0.48

10) To find T ', the value for R' is inserted into the following equation:

$$T' = 0.520 - \frac{1.815 * 0.48}{\sqrt{18}} = 0.315$$

11) T ' (0.315) is less than T (0.704), therefore the median baseline pollution loading is not exceeded.

2.2.2 Method 2 (Wilcoxon-Mann-Whitney Test) (See Figure 3.2b)

Instructions for the Wilcoxon-Mann-Whitney test are given in Conover (1980), cited in Figure 3.2b.

1) When using both baseline and monitoring data, n = 18 and m = 18.

2) The baseline and monitoring observations are listed in order of collection, and ranked as follows:

Baseline Observations		Monitoring Observations	
(lbs/day)	(Ranking)	(lbs/day)	(Ranking)
0.030	2.5	0.530	20
0.005	1	0.040	6
1.915	34	1.040	30
0.673	25	0.033	4
0.064	10	0.030	2.5
0.063	9	0.230	13
0.607	23	0.710	27
0.553	21	0.240	14
0.286	15	0.390	17
0.106	11	0.830	28
0.406	18	3.050	36
1.447	32	0.580	22
0.900	29	1.180	31
0.040	5	0.510	19
2.770	35	0.046	8
1.803	33	0.690	26
0.160	12	0.630	24
0.045	7	0.370	16

The value of 0.030 was obtained for more than one observation. The ranking displayed is the average of 2 and 3 (2.5).

3) The sum of the 18 baseline ranks (S_n) = 322.5.

4) From the table in section 1.2.2 of this appendix, the critical value (C) for 18 baseline and 18 monitoring observations is 238.

5) S_n (322.5) is greater than the critical value C (238). Therefore, according to the Wilcoxon-Mann-Whitney test, the monitoring observations did not exceed the baseline pollution loading.

3.0 Example 3

Assume 12 baseline flow and iron concentrations are collected by sampling once per month for a year. Likewise, 12 flow and iron monitoring observations are obtained by sampling once per month for a period of one year. In order to determine whether baseline pollution loading has been exceeded, both Methods 1 and 2 were used. For all calculations in Example 3, assume the following flows (in gpm) and iron concentrations (in mg/L).

Flow

Baseline	5.0	12.0	15.0	34.0	21.0	11.0	16.0	9.0	10.0	11.0	9.0	13.0
Monitoring	7.0	11.0	17.0	29.0	22.0	12.0	13.0	14.0	10.0	12.0	11.0	9.0

Iron Concentration

Baseline	11.4	8.2	6.0	11.1	6.4	10.3	12.1	14.2	6.1	8.3	10.0	13.5
Monitoring	12.3	13.5	9.8	7.9	5.8	7.5	8.2	9.3	8.4	12.5	14.1	15.3

Because there are three baseline concentrations (6.0 mg/L, 6.4 mg/L and 6.1 mg/L) below the Subpart C effluent limit for iron (7.0 mg/L), two separate sets of loading results are calculated. The first calculates iron loading using all the unmodified concentrations, and the following standard equation:

Load (in lbs/day) = Flow (in gpm) * Concentration (in mg/L) * 0.01202.

The resulting iron loads are given below:

Iron Load (using unmodified concentrations)

Baseline	0.69	1.18	1.08	4.54	1.62	1.36	2.33	1.54	0.73	1.10	1.08	2.11
Monitoring	1.03	1.78	2.00	2.75	1.53	1.08	1.28	1.57	1.01	1.80	1.86	1.66

The second set of calculated iron loads are calculated after replacing the baseline iron concentration below 7.0 mg/L with 7.0 mg/L. This set is given below, with the three modified loads in bold:

Iron Load (using modified concentrations)

Baseline	0.69	1.18	1.26	4.54	1.76	1.36	2.33	1.54	0.84	1.10	1.08	2.11
Monitoring	1.03	1.78	2.00	2.75	1.53	1.08	1.28	1.57	1.01	1.80	1.86	1.66

3.1 Single Observation Trigger

3.1.1 Method 1 (See Figure 3.2a):

1) Twelve baseline observations were collected, therefore n = 12.

2) The modified baseline observations were placed in sequential order from smallest to largest.

   [0.69, 0.84, 1.08, 1.10, 1.18, 1.26, 1.36, 1.54, 1.76, 2.11, 2.33, 4.54]

3) The number of observations, n, is less then 16, therefore the Single Observation Trigger (L) equals $x_{(12)}$, (the maximum) = 4.54.

4) All monitoring observations are less than 4.54, therefore the Single Observation Trigger (L) (4.54) was not exceeded.

3.1.2 Method 2 (See Figure 3.2b):

1) Twelve is an even number, therefore the median of the modified baseline observations is:

   $M = 0.5 * (x_{(6)} + x_{(7)}) = 1.31.$

   In order to determine M_1, calculate the median of the subset ranging from $x_{(7)}$ to $x_{(12)}$. Because 12 - 6 = 6 is even, $M_1 = 0.5 * (x_{(9)} + x_{(10)})$

   $M_1 = 0.5 * (1.76 + 2.11) = 1.935$

   Because M_1 is needed to calculate R, which must be based on unmodified concentrations, M_1 must also be calculated based on unmodified concentrations. The unmodified loads are ordered sequentially below:

   [0.69, 0.73, 1.08, 1.08, 1.10, 1.18, 1.36, 1.54, 1.62, 2.11, 2.33, 4.54]

   The median of unmodified baseline loads is:

   $M = 0.5 * (x_{(6)} + x_{(7)}) = 1.27.$

   The third quartile M_1 of the unmodified baseline loads is:

   $M_1 = 0.5 * (x_{(9)} + x_{(10)}) = 1.865.$

2) The first quartile M_{-1} of the unmodified baseline loads is:

$M_{-1} = 0.5 * (x_{(3)} + x_{(4)}) = 1.08.$

3) Using the values of M_{-1} and M_1 calculated using the unmodified baseline loads,
$R = 1.865 - 1.08 = 0.785.$

4) $L = M_1 + (3 * R) = 1.935 + (3 * 0.785) = 4.29$

4) All monitoring observations are less than 4.29, therefore the Single Observation Trigger (L) was not exceeded.

3.2 Annual Comparison

3.2.1 Method 1 (See Figure 3.2a)

1) Twelve is an even number, therefore the median of the modified baseline observations is:
$M = 0.5 * (x_{(6)} + x_{(7)}).$
$M = 0.5 * (1.26 + 1.36) = 1.31$

The following steps are needed to calculate R. Therefore, the unmodified baseline loads must be used. These load observations are listed in listed from sequential order from smallest to largest:
[0.69, 0.73, 1.08, 1.08, 1.10, 1.18, 1.36, 1.54, 1.62, 2.11, 2.33, 4.54]

In order to determine M_1, calculate the median of the subset ranging from $x_{(7)}$ to $x_{(12)}$. Because 12 - 6 = 6 is even, $M_1 = 0.5 * (x_{(9)} + x_{(10)})$
$M_1 = 0.5 * (1.62 + 2.11) = 1.865$

2) In order to determine M_{-1}, calculate the median of the subset ranging from $x_{(1)}$ to $x_{(6)}$. Because 6 is even, $M_{-1} = 0.5 * (x_{(3)} + x_{(4)})$
$M_{-1} = 0.5 * (1.08 + 1.08) = 1.08$

3) To calculate R, subtract M_{-1} from M_1.
$R = 1.865 - 1.08 = 0.785$

4) The calculated value for R is then substituted into the equation for T.

$$T = 1.31 + \frac{1.815 * 0.785}{\sqrt{12}} = 1.72$$

5) The following monitoring observations are ordered from smallest to largest.
[1.01, 1.03, 1.08, 1.28, 1.53, 1.57, 1.66, 1.78, 1.80, 1.86, 2.00, 2.75]

6) There are 12 monitoring observations, therefore m = 12.
The number of observations is even, therefore $M' = 0.5 * (x_{(6)} + x_{(7)})$
$M' = 0.5 * (1.57 + 1.66) = 1.615$
This holds true for M_1' and M_{-1}' as well.
$M_1' = 0.5 * (x_{(9)} + x_{(10)}) = 0.5 * (1.80 + 1.86) = 1.83$
$M_{-1}' = 0.5 * (x_{(3)} + x_{(4)}) = 0.5 * (1.08 + 1.28) = 1.18$

7) To calculate R, subtract M_{-1}' from M_1'
$R' = 1.83 - 1.18 = 0.65$

8) The calculated value for R' is then substituted in the equation for T'.

$$T' = 1.615 - \frac{1.815 * 0.65}{\sqrt{12}} = 1.274$$

9) T ' (1.274) is less than T (1.72), therefore the median baseline pollution loading was not exceeded.

3.2.2 Method 2 (Wilcoxon-Mann-Whitney Test) (See Figure 3.2b)

1) When using both baseline and monitoring data, n = 12 and m = 12

2) The modified baseline and monitoring observations are listed with their corresponding rankings.

Baseline Observations (lbs/day)	0.69	0.84	1.08	1.10	1.18	1.26	1.36	1.54	1.76	2.11	2.33	4.54
Baseline Rankings	1	2	5.5	7	8	9	11	13	16	21	22	24
Monitoring Observations (lbs/day)	1.01	1.03	1.08	1.28	1.53	1.57	1.66	1.78	1.80	1.86	2.00	2.75
Monitoring Rankings	3	4	5.5	10	12	14	15	17	18	19	20	23

Due to the fact that the value of 1.08 was obtained for two observations, an average ranking is used for this value. For 1.08, the average of 5 and 6 is 5.5.

3) The sum of the twelve baseline ranks (S_n) = 139.5.

4) From the table in section 1.2.2 of this appendix, the critical value (C) for 12 baseline and 12 monitoring observations is 99.

5) S_n (139.5) is greater than C (99). Therefore, according to the Wilcoxon-Mann-Whitney test, the monitoring observations did not exceed the baseline pollution loading.