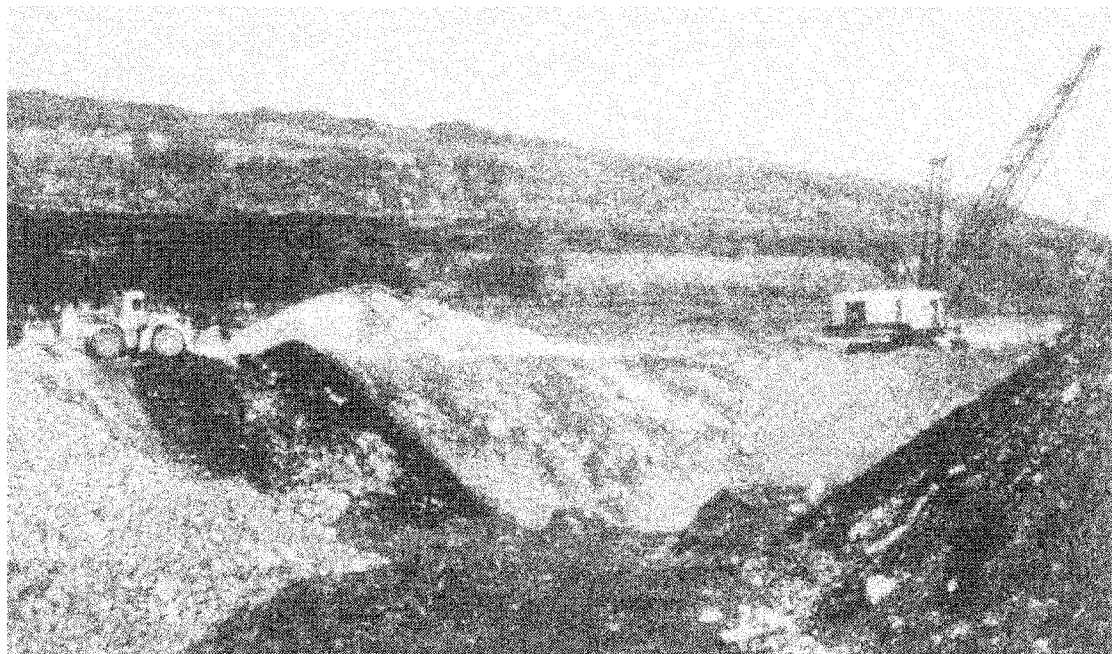# Characterization and Effectiveness of Remining Abandoned Coal Mines in Pennsylvania

By Jay W. Hawkins

UNITED STATES DEPARTMENT OF THE INTERIOR

UNITED STATES BUREAU OF MINES

U.S. Department of the Interior
Mission Statement

As the Nation's principal conservation agency, the Department of the Interior has responsibility for most of our nationally-owned public lands and natural resources. This includes fostering sound use of our land and water resources; protecting our fish, wildlife, and biological diversity; preserving the environmental and cultural values of our national parks and historical places; and providing for the enjoyment of life through outdoor recreation. The Department assesses our energy and mineral resources and works to ensure that their development is in the best interests of all our people by encouraging stewardship and citizen participation in their care. The Department also has a major responsibility for American Indian reservation communities and for people who live in island territories under U.S. administration.

Report of Investigations 9562

Characterization and Effectiveness of Remining Abandoned Coal Mines in Pennsylvania

By   Jay W. Hawkins

CONTENTS

ILLUSTRATIONS

ILLUSTRATIONS

TABLES

UNIT OF MEASURE ABBREVIATIONS USED IN THIS REPORT

Metric Units

g	gram	mg/L	milligram per liter
ha	hectare	mm	millimeter
kg/d	kilogram per day	t	metric ton
km	kilometer	t/ha	metric ton per hectare
L/min	liter per minute	t/yr	metric ton per year
m^3	cubic meter		

U.S. Customary Units

ft	foot	st	short ton
gpm	gallon(s) per minute	st/acre	short ton per acre
lb/d	pound(s) per day	yd^3	cubic yard

OTHER ABBREVIATIONS AND ACRONYMS USED IN THIS REPORT

AMD	acid mine drainage	NP	neutralization potential
BPJ	best professional judgment	NUPL	nonparametric upper prediction limits
CFR	Code of Federal Regulations	P	probability
COA	consent order and agreement	PADER	Pennsylvania Department of Environmental Resources
EPA	Environmental Protection Agency		
MW	Mann-Whitney U	SES	simulated effluent standards

Characterization and Effectiveness of Remining Abandoned Coal Mines in Pennsylvania

By Jay W. Hawkins[1]

ABSTRACT

Under an approved remining program, mine operators can remine abandoned coal mines without assuming legal responsibility for treatment of the previously degraded water, as long as the discharging waters are not further degraded and other regulatory requirements are satisfied. A U.S. Bureau of Mines review of 105 remining permits in Pennsylvania indicates that remining results in substantial reclamation of abandoned mine lands, utilization of significant quantities of coal, and reduction of contaminant loads (acidity and iron) from degraded mine drainage discharges. Normality tests performed on the water quality and flow data indicate generally nonnormal distributions and extreme right-skewness, tending toward lower values. The water quality of underground coal mines was observed to be more highly degraded in terms of acidity, iron, and sulfate than that of surface coal mines. The optimum baseline sampling scenario is 12 months in duration at a frequency of one sample per month. Analysis of water quality and flow rates before and after remining indicates that a majority of the mines exhibited either no change or a significant decrease in pollution rate because of remining. The discharge flow rate was the dominant controlling factor when the post-remining contaminant load was significantly better or worse than the baseline (pre-remining) load.

---

[1]Hydrologist, Pittsburgh Research Center, U.S. Bureau of Mines, Pittsburgh, PA.

INTRODUCTION

Remining operations have the potential to reclaim large areas of abandoned surface and underground coal mines of the Appalachian region without the use of tax-generated funds. Surface mining prior to the enactment of the Surface Mining Control and Reclamation Act of 1977 left large areas of abandoned exposed highwalls, open pits, and unregraded spoil piles (figure 1). It has been estimated that over 28,962 km (18,000 miles) of abandoned highwalls currently exist in the Appalachian region. Of the land disturbed by coal mining between 1930 and 1971, roughly 30% has been reclaimed. Estimates indicate that approximately 91 billion t (100 billion st) of recoverable coal exists within 180 m (590 ft) of abandoned highwalls (Lineberry and others, 1990). Previous underground mining has likewise left vast areas of abandoned mine workings, related surface subsidence features, and open mine entries (figure 2). Regulatory agencies of States in the Appalachian coal region where remining is not currently practiced may be inclined to start and promote remining programs if such programs can be shown to be successful in terms of enhanced coal recovery, reclamation of abandoned mine lands, and the reduction of (or no net increase in) degraded mine drainage. Mine operators may also be more inclined to enter into a remining project with the knowledge that the potential of incurring liability for long-term treatment of mine waters from prior mining activities is low. Furthermore, an in-depth analysis of remining may lead to improvements to existing programs.

"Remining," as the term is used in this report, is the surface mining of abandoned surface and/or underground mines that originally created and continue to discharge effluent water that fails to meet the applicable effluent standards for acidity and iron. Others have used the term "remining" to refer to the mining of abandoned surface or underground mines regardless of preexisting water quality. Under an approved remining program, an operator can legally mine such sites without assuming responsibility for treatment of the previously degraded water, as long as the discharging waters are not further degraded by the operation (technically, a slight water quality improvement may be required). If the water is additionally degraded because of the remining operation, the level of treatment required is based on pre-remining contaminant load levels and not on the legislatively promulgated effluent standards. In order to establish site-specific pre-remining contaminant load levels (baseline loading rates) and to ascertain water quality changes caused by remining,

Figure 1

Example of abandoned unreclaimed surface mine exhibiting flooded pit, exposed highwall, and sparsely vegetated "dead" spoil.

the mine operator must collect a series of pre-remining discharge water samples as well as discharge flow measurements. Loading of a given contaminant is determined by multiplying the discharge flow rate by the contaminant concentration. The strength of the pollution abatement plan and the economics of conventionally treating the water are also factored into the final baseline loading rates. To receive a remining permit, an operator must demonstrate that there is a potential to improve the water quality. Statistical analyses, primarily types of exploratory data analyses, are used to determine whether the post-remining discharges have been further degraded from baseline levels. If, after remining, the contaminant loading rates are within or below the established limits, based on the baseline loading rates, and all other post-remining and reclamation physical and temporal requirements are satisfied, discharge monitoring ceases and the operator's bonds are released.

Without a remining program, the mine operator would be liable for treatment in perpetuity for all discharges hydrologically connected with the site that failed to meet the applicable statutory effluent standards during and after reclamation. Even if the discharges were created by previous mining operations totally unrelated to the present operation and the water quality was improved by the remining, but remained below effluent standards, the operator would still be liable for perpetual treatment. For these reasons, prior to the initiation of a remining program, mine operators have avoided previously mined sites with existing contaminated discharges and minable coal reserves. However, in order to qualify for remining relief from statutory effluent standards, the operator must agree to perform some amount of reclamation of the previously abandoned mine lands and must illustrate that conventional treatment of the discharges to meet statutory effluent standards is cost prohibitive. Exactly how much reclamation is required is discretionary on the part of the State regulatory agency. Generally, spoil piles have to be regraded and revegetated to blend in with the existing topography; surface water impoundments have to be filled in; highwalls have to be eliminated, in some cases reclaimed to premining conditions; and exposed mine entries have to be sealed and

Figure 2

Example of abandoned underground mine exhibiting exposed mine entry and mine drainage

subsequently buried. Areas to be reclaimed during the operation are usually constrained to those within the permit boundary.

The objectives of this study include a review of the overall scope of remining in Pennsylvania, characterization of the mine water quality before and after remining, determination of the optimum pre-remining sampling scheme (in terms of sampling frequency and duration) to characterize the water quality, determination of the over-all effectiveness of remining in reducing the contaminant load, and identification of pollution abatement technologies that most effectively reduce or eliminate the contaminant load. This work is in concert with the U.S. Bureau of Mines (USBM) mission to ensure that the Nation has a dependable supply of minerals with minimal environmental impact.

BACKGROUND INFORMATION AND DATA COLLECTION

This study was based on data collected from the Pennsylvania remining program because Pennsylvania has a fully operational, well-documented, and time-tested remining program. Pennsylvania has been issuing remining permits for over 10 years, and other Appalachian States have begun the issuance of remining permits much more recently. Remining programs in many of these States are in their infancy or still in the formulation stages. Results of this project are based on the Pennsylvania program in terms of types of remining and abatement, stratigraphic and geographic scope, basic hydrologic information, and operational costs, which come from the information contained in the remining permit files. However, an in-depth analysis of the water quality and other quantitative data of these sites is the main thrust of this project.

The Pennsylvania Department of Environmental Resources (PADER), Bureau of Mining and Reclamation, has been issuing remining permits in the bituminous coal fields since 1983. However, the legislated remining program was not approved until a few years later. The enactment of Act 158 of 1984, which was an amendment to the State Surface Mining Conservation and Reclamation Act, formally permitted remining in Pennsylvania (Pennsylvania Bulletin, 1985). The remining program in Pennsylvania (Act 158, Subchapter F and G programs) was introduced by use of a permit module (module 26) and a series of standard special conditions by early 1986. These changes to the Pennsylvania coal mining regulations required the concurrent approval of the U.S. Office of Surface Mining Reclamation and Enforcement and the U.S. Environmental Protection Agency (EPA), in accordance with the Federal Surface Mining Control and Reclamation Act. The EPA concluded that effluent guidelines for discharges from active coal mining operations (40 CFR, part 434) do not apply to preexisting discharges that are not physically encountered by a remining operation. To permit these changes to Pennsylvania law, the EPA determined that preexisting unencountered discharges required a case-by-case establishment of effluent standards using best professional judgment (BPJ) analyses under the provisions of section 402(a)(1) of the Federal Clean Water Act (Pennsylvania Bulletin, 1985). Determination of the BPJ analyses for remining situations is described by Kohlmann

Ruggiero Engineers (1986, 1990). Additionally, Phelps and Thomas (1986) determined the economic and technical feasibility of surface mine reclamation involved in the BPJ analysis. It should be noted that if the degraded discharges are physically encountered during mining, they must be treated to meet the standards specified by 25 PA. CODE, part 87.102, which is based on the effluent standards established by 40 CFR, part 434. Once these discharges are no longer physically encountered by the mining operation, the modified effluent standards, under the Subchapter F program, are reinstated. Reinstatement generally occurs during the reclamation stages of a remining operation. The statistical analysis of the water quality data to create the modified effluent standards using the BPJ analyses is performed by a computer program (REMINE) created for the Pennsylvania remining program (PADER and others, 1988).

Approximately 20 permits were issued prior to the introduction of the formal Subchapter F program and module 26 of the permit application. In these permits, remining provisions were incorporated by use of a consent order and agreement (COA); later, "special conditions" were appended to the permit. Most of the background information pertaining to the remining operation was contained in the COA. Under Subchapter F, most of the background information and data pertaining to remining are contained in the permit file, while provisions and obligations of the permittee are incorporated into the permit by use of the standard special conditions. As of March 1992, over 90 of the latter type of remining permit had been issued in the Pennsylvania bituminous district, bringing the total number of permits to well over 100.

For this project, mining, pollution abatement, reclamation, and hydrologic data from 105 remining permits in the bituminous region of central and western Pennsylvania were collected (figure 3). The majority of the sites that were reviewed had yet to be activated by remining or were still being actively mined. For 24 of the surveyed sites, mining has been completed (the sites have been backfilled to rough grade), and at least 1 year of post-remining water quality and flow data have been collected.

Figure 3

LEGEND

Bituminous coal fields

Scale, km
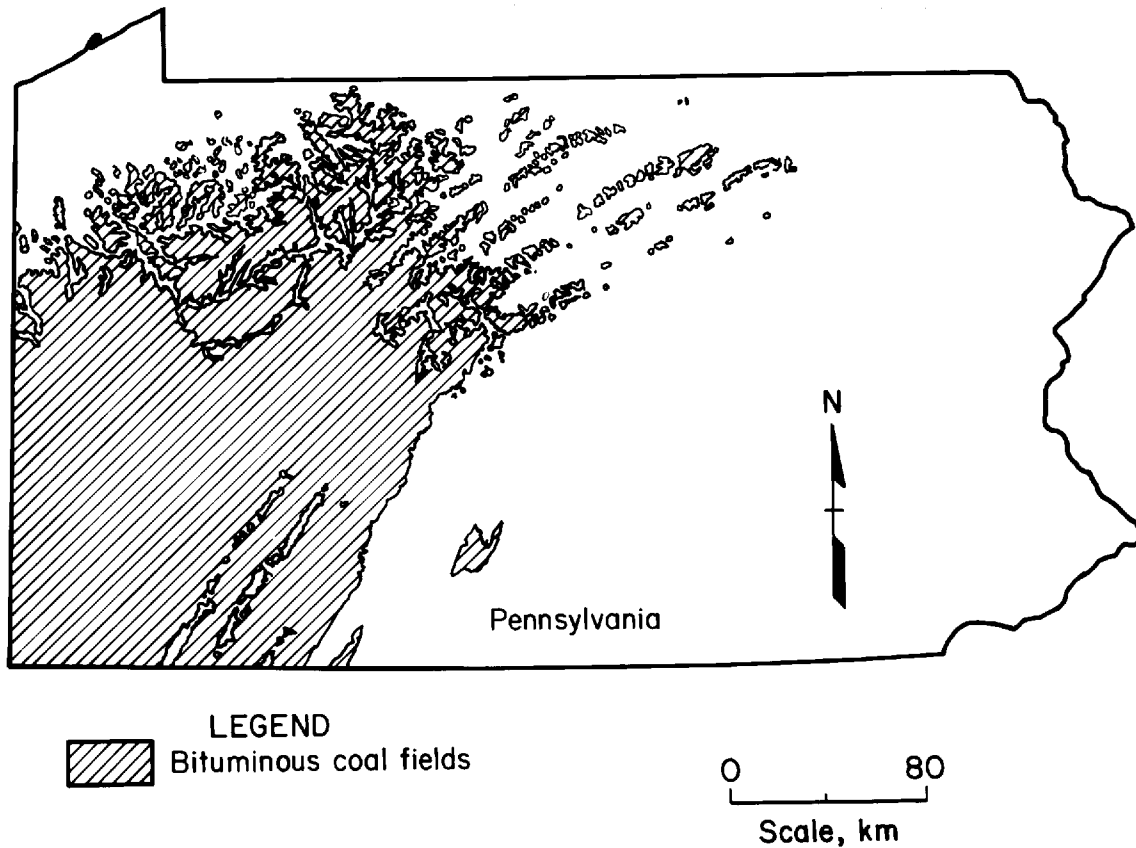
Study area within the bituminous coal fields.

WATER QUALITY STANDARDS FOR REMINING

The effluent standards for preexisting mine discharges on remining sites are based on the pre-remining water quality and flow rates. Remining effluent standards are set as baseline contaminant loading rates calculated by multiplying contaminant concentration by flow rate, which are reported in units of pounds of contaminant per day. This is in contrast to the usual contaminant effluent limits, which are in units of contaminant concentration (e.g., milligrams per liter), as set by EPA regulation in 40 CFR, part 434.30, and by Commonwealth of Pennsylvania regulation in 25 PA CODE, Chapter 87, part 87.102.

Sample Collection

Baseline contaminant loading rates are set during the re-mining permit application process using a temporally con-secutive series of pre-remining water quality samples along with measured discharge flow rates. Initially, a minimum of 6 monthly samples were required to perform these calculations, although 12 consecutive monthly samples were strongly recommended by the PADER. Some permits issued prior to 1986 under the Pennsylvania remining program had fewer than six samples because that was permitted at the time. The PADER now requires 12 consecutive months of data or at least samples collected from February through October. In Pennsylvania, sampling the period from February through October will usually record the highest and lowest loading rates (Smith, 1988). Calculation of baseline loading rates uses consecutive monthly sampling from an entire water year (Octo-ber 1 through September 30) or water years. Partial water-year data cannot be used, unless sampled from February though October. In theory, 12 consecutive monthly samples will include both dry and wet seasons in the background data set, which will more accurately characterize the preexisting discharges.

Sample Analysis

At a minimum, the water samples must be analyzed for con-centrations of alkalinity, acidity, total iron, total manganese,

aluminum, sulfate, total suspended solids, and pH. The contaminants are generally reported in units of milligrams per liter, except for pH, which is in standard units. Discharge flow rate, usually reported in gallons per minute, is gaged by various means including the use of weirs, flumes, cross-sectional area with a flow meter, and the bucket-stopwatch method. The weir appears to be the most common method used for mine discharges over 38 L/min (10 gpm), while the bucket-stopwatch method appears to be the main method for discharges with lower flow rates. For the remining permits, the loading rates in pounds per day are calculated by multiplying the flow and concentration data. For this study, the flow data were converted to liters per minute, and the loading data were converted to kilograms per day.

The pre-remining water quality data are analyzed using basic exploratory data analyses and nonparametric statistics. The results are presented in a tabular format containing the data range, the median, the first and third quartiles, the approximate 95% tolerance limits (depth of 32nds values or C spread), and the 95% confidence interval about the median for each of the regulated contaminants (figure 4). For additional information on how these values are calculated and what they represent, the reader is directed to Tukey (1976) and/or Velleman and Hoaglin (1981). These statistics become the site-specific tolerance limits within which loading rates will be regulated during and after remining.

Mechanisms that Trigger Treatment

Under the PADER system, there are four mechanisms by which treatment of a discharge can be triggered (initiated)

using this table. The first triggering method under this system requires a series of six consecutive samples to exceed the upper bound of the approximate 95% tolerance limits (item 4 on figure 4). During and after mining, discharge sampling is performed on a monthly basis until all reclamation performance bonds are released (generally Stage II bonds). If two consecutive samples exceed the upper bound of the approximate 95% tolerance limits for any of the specified contaminants, this immediately triggers weekly sampling of the discharge. If four consecutive weekly samples exceed the approximate 95% tolerance limits, then the operator must initiate treatment within 30 days. If two consecutive weekly samples drop below the 95% tolerance limits, then monthly monitoring resumes and treatment is not required at that time. The mechanism to suspend treatment is not well defined. No clear policy currently exists.[2]

Treatment can also be initiated (by the second triggering method) if statistical analysis of the data indicates that the median contaminant load during- or after-remining has been increased compared with the pre-remining median at the 5% significance level. This is determined by comparison of the 95% confidence interval about the median (figure 4, item 5) of the pre- and post-remining data. For this method, the median is calculated on a complete water-year basis (October 1 through September 30).

The third triggering method uses the same method of analyses. However, the median is determined for water-year periods 1 (October 1 through April 30) or 2 (May 1 through

[2]Michael Smith, PADER, personal communication.

Figure 4

Mine ID:	Mine Name:		Hydrologic Unit ID:			
Loading in pounds per day						
Parameter: Number of samples (N):			Flow (gpm) 43	Acidity 43	Iron 43	Sulfate 43
1. Range	Low:		3.00	0.07	0.00	28.27
	High:		42.00	1.01	1.41	214.56
2. Median			12.00	0.29	0.21	99.53
3. Quartiles	Low:		9.00	0.22	0.15	70.98
	High:		17.00	0.41	0.60	132.63
4. Approximate 95% tolerance limits	Low:		3.00	0.07	0.02	31.01
	High:		34.00	0.82	1.29	210.47
5. 95% Confidence int. about median	Low:		10.22	0.25	0.11	85.77
	High:		13.78	0.33	0.31	113.28

Example of summary table of baseline contaminant loads, as required by PADER. Units are those used for remining pernits by PADER.

September 30). Finally, treatment can be triggered if statistical analyses, including but not limited to the means and variances of the data, indicate that the difference between water years or water-year periods is significant at the 1% level (exceeds the 99% confidence level).

For any of the triggering mechanisms, if the mine operator can demonstrate to the PADER that the apparent increase in contaminant load is unrelated to the min-ing operation and is caused by factors beyond control of the remining operation (e.g., adjacent unrelated mining operations or an extreme storm event), treatment of the discharge will not be required.

Treatment standards are likewise based on the data contained in figure 4. The treated monthly discharge contaminant load average must be equal to or less than the pre-remining median (item 2 on figure 4) and is calculated based on the samples collected weekly. The instantaneous maximum contaminant load permitted is based on a "grab" sample and must be no greater than the upper quartile ("High" value of item 3 on figure 4). Both parameters must be reported monthly, at a minimum.

GENERAL SCOPE OF REMINING

Stratigraphically, PADER remining permits have been issued on 21 separate coal seams from the Pottsville Group (Mercer Coal, Lower Pennsylvanian age) through the base of the Dunkard Group (Waynesburg Coal, Lower Permian age). Not unexpectedly, the bulk of the permits have been issued on coal seams that have historically seen considerable surface and underground mining. Approximately 51% of the mining permits were issued to remine the Freeport and/or Kittanning Coal Seams (Allegheny Group). An additional 25% of the mining permits were issued to remine the Pittsburgh Coal Seam (Monongahela Group). Roughly 45% of the permits were for multiple-seam operations, with up to six separate seams and/or rider seams mined under one permit. Ten percent of the permits were issued to allow coal refuse reprocessing to remove the residual coal in abandoned gob (coal refuse) piles.

BASIC STATISTICS

Mine areas permitted range in size from 1.3 to 310 ha (3.1 to 766 acres), with an average of 45.7 ha (113 acres) (table 1). However, the actual abandoned mine area within the permitted boundary ranges from 0.4 to 160 ha (1 to 395 acres), with an average of 27.2 ha (67 acres) and a median of 17.2 ha (42.5 acres). The abandoned area expressed in the permit to be reclaimed during remining ranges from 0.2 to 71.2 (0.5 to 176 acres), with an average of 15.3 (38 acres) and a median of 9.3 ha (23 acres). The average abandoned area slated for reclamation is essentially the same for underground and surface mines, 14.3 and 13.8 ha (35.3 and 34.1 acres), respectively. Overlap associated with multiple-seam mining can cause slight differences between the total of the average abandoned and/or reclaimed mine area and the sum of underground and surface mine areas determined separately (table 1). The average surface mine area slated to be reclaimed by remining, 10.1 ha (25 acres), is substantially higher than the average underground mine area to be reclaimed, 5.6 ha (14 acres). This difference is most likely because the underground mines commonly yield

50% or less coal per unit area than surface mines, and accurate determination of the coal reserves for underground mines beforehand is extremely difficult. Mine operators are also probably less inclined to remine more of the underground mine workings because the amount of ground water stored in and moving through underground mines is commonly substantially greater than that of surface mines. This additional mine water increases mining costs and is perceived to represent a higher risk of incurring post-remining treatment. The percentage of abandoned mine land within a mine permit boundary to be reclaimed averages 69%, with a range of 4% to 105%. Multiple-seam mining allows the reclaimed percentage to exceed 100%, because abandoned mines on two or more seams can overlap and the percentage reclaimed is calculated based on the total area given in the permit.

Table 1.—Total, abandoned, and reclaimed areas, hectares

	Low	High	Mean	Median
Total permitted	1.3	310.0	45.7	30.5
Total abandoned	0.4	160.0	27.2	17.2
Total reclaimed	0.2	71.2	15.3	9.3
Surface mine abandoned	0.0	71.0	13.8	6.3
Underground mine abandoned . .	0.0	156.0	14.3	5.0
Reclaimed surface mine	0.0	62.7	10.1	5.0
Reclaimed underground mine . .	0.0	39.0	5.6	2.8

The amount of coal recovered by these types of operations varies widely (two orders of magnitude) from 11,794 to over 2,177,280 t (13,000 to over 2,400,000 st). The amount of coal recovered is extremely important, because this coal represents a mineral resource that might not otherwise be recovered, without relief of treatment liability for pre-remining mine drainage discharges. Without relief under the remining program, most mine operators would be unwilling to take the risk of having to provide perpetu-al mine drainage treatment, and the coal resource of abandoned sites would never be utilized. Furthermore, if a site is reclaimed through the

abandoned mine land program, reopening it at a later date to extract the coal may not be economically feasible. The average amount of coal recovered per mine is over 317,520 t (350,000 st). However, the median, 186,998 t (206,127 st), is a better indicator of the central tendency of the coal tonnages, because the data are nonnormally distributed at the 99% confidence level and are strongly skewed toward the lower values (right) using the chi-square test. Figure 5 is a histogram illustrating the distribution of these data.

The number of point-source discharges prior to mining at the 24 mines studied ranges from 1 to 30. The average number of discharges is four. In many of the permits, several discharges have been combined to form hydrologic units for simplicity of data analysis and sampling cost savings, and because the location and number of discharges commonly change after the site is mined. A hydrologic unit is a portion of a surface mine comprising discharges that are physically and/or hydrologically connected. If a discharge relocates after remining, use of hydrologic units allow comparison to known baseline conditions. Hydrologic unit boundaries are usually defined prior to permit issuance, although they may be modified with additional geologic and hydrologic information obtained as the site is remined. Combining of discharges can be performed physically, where they are actually routed to a common collection and monitoring point, or it can be performed mathematically by summing the raw data collected from each discharge point of a hydrologic unit prior to analyses.

Median acidity loading rates before remining range from less than 0.5 to over 4,880 kg/d (1 to over 10,760 lb/d), with an average of 165 kg/d (363.8 lb/d) and a median of 20.3 kg/d (44.8 lb/d). The sites with the lowest loading rates are generally small abandoned surface mines, while the highest loading rates have been recorded at a really extensive underground mines. Both high flow rates and generally high concentrations associated with underground mines are the cause of these higher loading rates. Median iron loading rates before remining ranged from 0.004 to over 816.5 kg/d (0.01 to over 1,800 lb/d), with an average of 17.2 kg/d (37.9 lb/d) and a median of 0.54 kg/d (1.2 lb/d). Like the acidity loading rates, the higher iron loads are directly related to the higher flows and concentrations of underground mine discharges. Similar trends were noted for pre-remining sulfate loads. Although sulfate is not an effluent standard parameter, it serves as an conservative indicator of the geochemical reactions taking place within the remined site. Sulfate loading rates range from 0.68 to over 10,342 kg/d (1.5 to over 22,800 lb/d). The median sulfate load is 73.6 kg/d (162.3 lb/d). Sulfate concentrations and loading rates may indicate, along with other data, the reason for success or failure of remining to reduce the concentration or contaminant load, because the sulfate ion is released in the formation

Figure 5
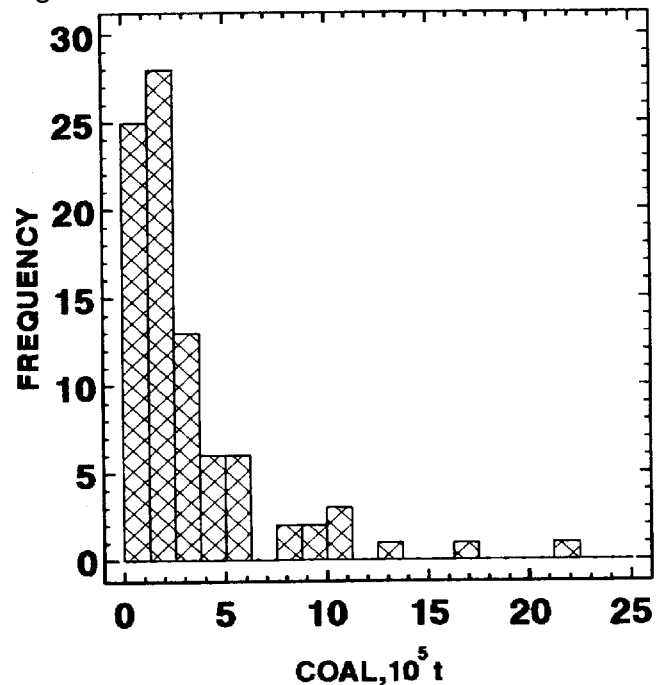
Frequency histogram of coal tonnages to be recovered by remining operations.

of acid mine drainage (AMD). The concentration of sulfate in the discharging water is indicative of the rate of AMD production.

POLLUTION ABATEMENT

To obtain a permit to remine, an operator must outline what actions will be taken in an attempt to abate or reduce the preexisting contaminant load. This information is included in the permit application abatement plan. The PADER fully recognizes that in most cases the pollution will not be completely eliminated, but the operator must nevertheless define what abatement techniques will be instituted in an attempt to reduce the severity of the degradation. The Pennsylvania remining program breaks down abatement techniques into eight discrete categories: regrading of abandoned surface mine spoils ("dead" spoils), underground mine daylighting (surface mining of the remaining coal by the removal of the overburden), revegetation, addition of alkaline material brought in from off site, special handling of acid-producing spoil materials, hydrologic control of ground and/or surface water, sewage sludge application, and all other remining types.

Abatement Techniques

The abatement plans of most of the operations (90%) included more than one technique to be employed during mining and reclamation. The most common forms of abatement were spoil regrading, underground mine daylighting, and revegetation; each type was listed for 70% of the permits. This was anticipated, because these techniques would usually be part of the standard mining or reclamation process whether or not the permits were issued for remining sites. Alkaline addition (34%), special handling of spoil materials (31%), and hydrologic control measures (29%) were also listed as abatement procedures for a significant number of permits. Because these techniques entail actions that are in addition to the normal surface mining procedures, it was not unexpected that they were listed for only one-third of the permits. They entail additional cost, time, and effort, and therefore are used only where the other abatement measures inadequately address the permit requirements for an abatement plan. None of the permits listed sewage sludge application as an abatement technique, and 5% listed "other." The lack of sewage sludge disposal as an abatement technique may be related to the considerable public opposition to this method, which usually prolongs and adds cost to the permitting process, as well as the reluctance of mine operators to employ a technique little known to them.

When regrading of dead spoils is part of the abatement plan, cubic meters or cubic yards of spoil regraded is a gage of the amount of abatement work to be performed. Dead spoil is overburden material that the original mining operation did not regrade or revegetate. The remining operator will generally need to transport or regrade this material while gaining little or no coal recovery. Data on spoil regrading was available for 54 of the permits. The amount to be regraded ranged from 3,823 to over 2,293,800 m³ (5,000 to over 3,000,000 yd³), with an average of 356,304 m³ (466,000 yd³).

As stated above, roughly 34% (34) of the 105 sites listed alkaline addition as part of the abatement plan. The rationale for alkaline addition is that it will prevent or abate the formation of AMD or neutralize AMD that has already formed. Carbonate-rich rocks (limestone or dolostone) were the most commonly listed alkaline additives (35%), when the type of material to be used was stated. Carbonate-rich rock is commonly used for alkaline addition because of its widespread availability and relatively low cost. Thirty-eight percent of the permits did not specify the type of alkaline material. The remaining 27% of the permits listed a variety of materials, such as alkaline coal ash or hydrated lime ($Ca(OH)_2$). A few of the operators (24%) had the material chemically tested to determine the neutralization potential (NP). The NP values cited by these operators ranged from 1.75 to 950 g of calcium carbonate equivalent per kilogram of material. The amount of alkaline material added per acre was listed in all 34 permits. The application rate ranged from 5.7 to 60,524 t/ha (2.5 to 27,000 st/acre). The site where the application rate is 60,524 t/ha is also a fly ash and bottom ash disposal permit, and the NP value is 58, which is somewhat low (5.8% calcium carbonate equivalent). The application rate for the majority of the sites (23) was below 112.2 t/ha (50 st/acre).

Theoretically, the location of the alkaline material placement within the mine backfill will determine the effectiveness of this practice in reducing the AMD contaminant load. The practice of "liming the pit floor" was used in 18 out of 34 (53%) of the permitted sites. The alkaline material was placed in the backfill, either intermingled or placed in discrete lifts, in six (18%) of the sites. On one site, the material was to be placed on top of the spoil and below the soil horizon. Three sites used the alkaline material as a soil amendment, in part to promote plant growth. Eight permits did not specify where the alkaline material was to be placed. A few permits listed more than one placement location.

Of the alkaline addition sites, eight also employed the abatement practice of regrading of dead spoils. Thirteen sites had alkaline addition in conjunction with underground mine daylighting. Eleven sites listed both regrading of dead spoils and underground mine daylighting along with alkaline addition. For the remaining two sites, alkaline addition was performed in conjunction with other abatement practices. Most of the alkaline addition sites (31) were for coals of the Monongahela and Allegheny Groups. This is not necessarily because these coals are any more prone to AMD production, but because these are the most commonly mined coals in the bituminous region of Pennsylvania.

Abatement Costs

The estimated costs of implementing the abatement procedures, as submitted by the mine operators or their permitting consultants, can be considerable. The costs range from $0, when the operator considers the abatement procedure to be part of the normal mining costs, to near $4 million, when a considerable amount of work above normal surface mining practices is required. The average cost is near $340,000. The average abatement cost per hectare is $25,723 ($10,410 per acre), with a range of $0 to $256,980 per hectare ($0 to $104,000 per acre). With reclamation bonds seldom exceeding $12,355 per hectare ($5,000 per acre), the higher range values may indicate that the operator is being unrealistic or that the bonding rate is inadequate to insure reclamation of remining sites.

Where the estimated abatement cost appears high, based on the author's experience, it may have been somewhat inflated in order to strongly illustrate to the PADER why the operator should be given relief from treating the existing discharges to the mandated effluent standards. This illustration is required as part of the remining program to show that a significant amount of reclamation is being achieved. Of the 89 permits in which abatement cost per metric ton of coal was determined, 24 listed abatement costs above $1.36 per metric ton ($1.50 per short ton) of coal produced. The highest value given was $7.73 per metric ton ($8.52 per short ton), which appears to be greatly inflated.

One of the mechanisms to illustrate why the operator should not be held accountable for treating the preremining discharges to the State effluent standards (25 PA. CODE, part 87.102), or somewhere between effluent standards and baseline background levels, is to show that the projected treatment costs would be prohibitive. This is done on an incremental basis of 1-, 5-, and 50-year treatment costs. The costs include treatment facility construction, materials, electricity, chemicals, maintenance, and sludge disposal costs. Table 2 lists the data from treatment cost projections. The costs range from a low of $11,044 for a single year of treatment to over $49 million for 50 years of treatment. The cost of treatment for 50 years per metric ton of coal ranges over three orders of magnitude,

$0.17 to $870.88 ($0.19 to $959.96 per short ton) with an average of $38.37 ($42.29 per short ton), or $0.74 per metric ton ($0.82 per short ton) per year. Normality testing of these values illustrates that they are nonnormally distributed and are strongly skewed to the right. The chi-square test illustrates that the nonnormal distribution is significant at the 99% confidence level. Therefore, the median, $5.90 per metric ton ($6.50 per short ton), is a better indicator of the central tendency of the data. These data were available for 86 of the 105 sites reviewed. Nine sites (10%) listed the 50-year treatment cost per metric ton of coal below $1.36 ($1.50 per short ton), and 37 sites (43%) listed a treatment cost that exceeded $9.07 per metric ton ($10.00 per short ton) of coal. The fact that roughly 90% of the sites had projected treatment costs exceeding $1.36 per metric ton ($1.50 per short ton) of coal indicates that for a vast majority of the cases, it is not an economically viable option for operators to assume long-term treatment liability, if they would have to meet State effluent standards.

Table 2.—Total projected water treatment costs

	Year 1	Year 5	Year 50
High	$1,009,569	$4,922,672	$49,069,007
Low	11,044	19,702	117,103
Average ...	121,618	489,056	4,545,035

WATER QUALITY

ANALYSIS OF DATA

In Pennsylvania, pre-remining water quality data are used to determine the baseline loading effluent standards that the operator will be held to during and after remining. Prior to remining, the mine operator must collect and chemically analyze a temporally consecutive series of mine discharge water samples and measure the discharge flow rate. The baseline contaminant loading rates are established using these data. These data are submitted as a required part of the remining permit application, as shown in figure 4. Statistical analyses, primarily exploratory data analysis (schematic summary), are employed to ascertain if the post-remining discharges have been degraded relative to baseline conditions. The statistical analysis is performed by using a computer program (REMINE) especially developed for the Pennsylvania remining program (PADER and others, 1988). If the contaminant loads are below limits based on baseline contaminant loads, the remining operator is released of liability at the completion of all other reclamation and time requirements.

To adequately evaluate the effectiveness of remining, a basic understanding of the water quality and discharge flow rate characteristics is required. Success of a remining operation is defined primarily by the lack of additional mine water

contamination (technically, a decrease in the contaminant load is required) and by the reclamation of abandoned mine lands. The amount of reclamation achieved by the operation is easily quantifiable in terms of abandoned mine areas regraded and revegetated, linear meters of highwall eliminated, or area of underground mines daylighted. However, the determination of changes in water quality and/or flow rate is much more difficult to quantify. In order to accurately assess changes caused by remining, trends and characteristics of the pre- and post-remining water quality and flow rates must be evaluated in an unbiased manner.

RELATED STUDIES

The majority of previous studies pertaining to analyses of ground water quality data are unrelated to coal mining or remining. This section was written to acquaint the reader with characteristics of hydrologic data from acid-producing coal mines using normality testing (skewness and chi-square), exploratory data analyses (notched box-and-whisker plots), and ranked correlation coefficient determinations (Spearman's rank correlation). The author fully acknowledges that there are numerous other testing techniques and methodologies applicable to these data that were not used here, including but not limited to seasonality and serial dependence tests.

There are very few studies regarding the environmental impacts of remining of coal mines. Smith (1988) and Hornberger and others (1990) analyzed the effects of sea-sonal variations on pre-remining flow rates, acidity concentrations, and acidity loads of three mine discharges in Pennsylvania. The discharges were located in three different hydrogeologic settings and exhibited distinctly different characteristics. Discharge characteristics were categorized as (1) "high flow - low concentration" or "low flow - high concentration," (2) "steady or damped" response (a slight or a delayed and subdued change in concentration in response to flow changes), and (3) "slugger" (discharge flow increases were not accompanied by any significant change in acidity concentration). Discharge flow rate was observed to dominate the acidity loading determinations; therefore, a strong positive correlation between flow and load is anticipated. The research indicated that sampling to determine contaminant loading rates should be of adequate duration and consistent frequency to accurately characterize both high and low flow periods of a discharge, not overemphasizing any one period (Smith, 1988).

Hornberger and others (1990) discussed spatial distribution of AMD in Pennsylvania and temporal variation patterns in different types of AMD discharges. They determined that 78% of the total amount of AMD was produced by abandoned or inactive coal mines. Underground mines accounted for approximately half of the sources, but contributed more than half of the AMD produced. They concluded that any effort to characterize an AMD discharge must take into consideration the common variability in flow and quality.

Helsel (1983) analyzed streams from mined and unmined watersheds in eastern Ohio to determine the influence of mine and rock type on water quality. He determined that overburden lithology influenced several water quality constituents. The water quality data exhibited neither a normal nor lognormal distribution. The effects of mine and rock type on water quality were more adequately shown using analysis of the ranks of the data, rather than analysis of the actual data values.

Previous research has indicated that under many non-mining-related circumstances, hydrologic data are generally nonnormally distributed (Berryman and others, 1988). Researchers have observed that water quality data tend to be asymmetric and skewed right (Hirsch and Slack, 1984; Montgomery and others, 1987). Given these nonnormal tendencies, the use of parametric statistics, which are sensitive to assumptions of normality, may lead to erroneous determinations of degradation or nondegradation. Montgomery and others (1987) and Helsel (1987) suggested using nonparametric statistical methods for nonnormally distributed data or performing some form of data transformation to approximate a normal distribution prior to statistical analyses. Logarithmic transformation will commonly eliminate skewness and asymmetry, transforming these data into an approximate normal distribution (Harris and others, 1987; Norcliffe, 1977).

There are numerous statistical methods for the determination of data normality. Harris and others (1987) stated that the skewness test is the best for ground water quality parameters and that the commonly used chi-square goodness-of-fit test does not work as well for these types of data. However, when the sample size is small (under 24), all of the tests for the assumption of normality begin to lose statistical validity (Montgomery and others, 1987).

DISCUSSION OF STATISTICAL METHODS

This section presents the results of univariate and bivariate statistical analyses of data from 57 mine discharges emanating from 24 coal remining operations in the bituminous coal fields in western Pennsylvania (figure 6). These 24 remining operations were selected from the larger group of 105 potential sites. The remining operations that were selected possessed sufficient post-remining hydrologic data (a minimum of 1 year, dating from rough backfilling) to permit direct comparison with the baseline data.

Of the hydrologic data obtained from the remining permits, only acidity, total iron, sulfate, and flow rate are discussed here. Under present Pennsylvania remining regulations, effluent standards for acidity and iron loading rates are mandated for all permits, regardless of the actual concentration of each parameter with respect to the legislatively mandated effluent standards. Sulfate, although not a regulated contaminant for surface water, was included in these analyses because it is generally a conservative indicator of AMD production. Increases in the sulfate content of mine water in the Appalachian coal mining region indicate an acceleration of metal sulfide, primarily iron disulfide (pyrite), oxidation. Sulfate ions (SO_4^{2-}) are released as a result of this reaction. Sulfate concentration is little affected by geochemical changes to the mine water (e.g., pH changes and increases in acidity and dissolved oxygen), and sulfate remains in solution to relatively high levels, governed primarily by the calcium concentration and the solubility of gypsum ($CaSO_4 \cdot 2H_2O$).

Figure 6
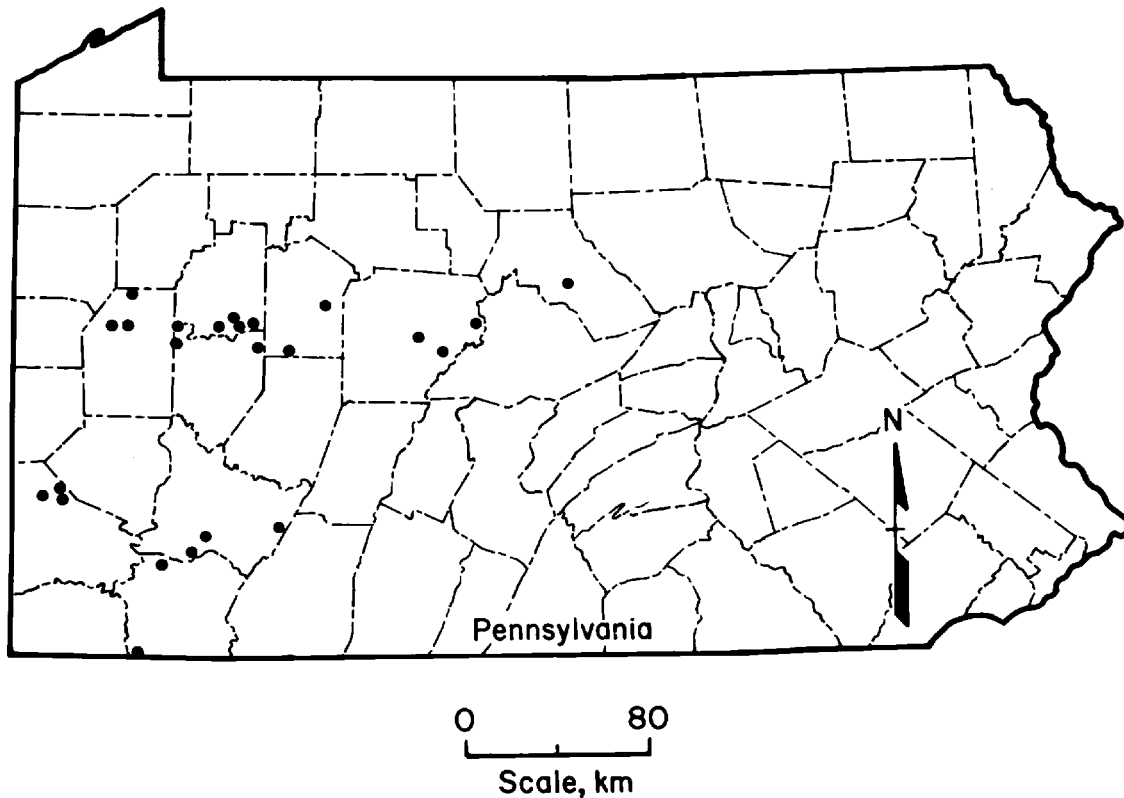
Location of 24 study sites used in comparison of pre- and post-remining data.

AMD is created when metal sulfide minerals (commonly pyrite) oxidize and the oxidation products are subsequently mobilized. Ground water serves as the transport medium of these oxidation products. Recharge events tend to "flush out" the oxidation products as the wetting front moves through the unsaturated portion of the spoil. When coal is surface mined, the overburden material is broken up into particles ranging in size from clay (<0.002 mm) to boulders (>256 mm). This fragmentation greatly increases the rock surface area, exposing additional pyritic minerals to atmospheric oxygen and iron-oxidizing bacteria. This promotes a state of geochemical flux for a period of time after mining. Based on subsidence observations and extensive aquifer testing, mine spoil continues to undergo considerable physical changes caused by compaction, shifting, and piping by ground water for at least 30 months after reclamation (Aljoe and Hawkins, 1992). Spoil continues to physically change well beyond these initial 30 months, but does so at a much reduced rate. All of these physical processes directly affect the hydraulic properties of the spoil aquifer.

The number of mine discharges sampled at each of the 24 sites ranged from 1 to 5, totaling 57. The pre- and post-remining data were analyzed as separate data sets because of significant physical and geochemical changes that can occur to the mine spoil aquifer during mining and subsequent reclamation. The baseline sampling period ranged from 3 to 42 months with the collection of 3 to 38 samples. All but two sites had a minimum of six preremining samples. The average pre-remining sample set contained 17 samples. The post-remining sampling period ranged from 12 to 65 months with 7 to 71 samples collected, with an average of 30 samples. Table 3 summarizes the median concentrations and flow rates for the 24 sites. Appendix A is a table summarizing the loading rates for these sites. For portions of the statistical analyses, the data were further differentiated into underground mine and surface mine discharges to ascertain potential differences between pre- and post-remining data based on these mine types.

TESTS FOR NORMALITY

Discharge flow rate, contaminant concentration, and loading rate data before and after remining were tested for normality using the skewness test and the chi-square

goodness-of-fit test. The results of these analyses are summarized in table 4. The data were tested for normality at the 5% significance level, and form of skewness (left or right) was determined.

Table 3.—Raw remining data[1]

Site			Pre-remining					Post-remining		
	n	Flow, L/min	Acidity, mg/L	Iron, mg/L	Sulfate, mg/L	n	Flow, L/min	Acidity, mg/L	Iron, mg/L	Sulfate, mg/L
1 ...	17	8	19	0.2	339	13	4	30	0.2	420
2 ...	21	1,181	321	24.1	814	41	1,283	261	20.5	852
3 ...	22	23	511	62.0	1,132	14	4	512	36.0	1,089
4 ...	10	144	43	2.9	92	45	155	162	21.9	602
5 ...	38	193	18	0.1	151	19	182	11	0.1	118
6 ...	21	110	143	2.7	732	16	140	128	2.3	722
7 ...	31	469	1,020	11.3	1,077	11	466	850	12.8	1,087
8 ...	10	204	777	99.3	1,695	40	117	294	37.6	2,189
9 ...	4	261	1,447	58.1	2,671	16	95	742	54.7	2,230
10 ..	9	250	4	0.6	53	33	144	3	0.5	51
11 ..	6	280	302	10.9	991	17	38	262	6.7	882
12 ..	12	462	58	0.2	NA	63	140	23	0.4	326
13 ..	11	53	5	0.2	NA	56	30	10	0.2	649
14 ..	28	11	9	1.2	159	46	15	9	0.8	204
15 ..	9	19	136	3.5	236	24	8	299	1.6	876
16 ..	3	189	456	295.0	1,430	71	148	541	218.5	1,202
17 ..	26	64	80	1.0	515	43	45	2	1.7	690
18 ..	24	265	2	0.4	153	33	348	10	0.3	270
19 ..	18	34	16	0.2	74	36	42	83	1.1	446
20 ..	8	140	208	3.3	740	7	4	19	0.5	749
21 ..	16	42	90	2.5	231	21	61	90	2.6	379
22 ..	28	238	231	4.8	931	10	428	151	5.9	779
23 ..	8	344	89	4.6	253	25	220	127	9.8	374
24 ..	18	11	566	64.1	673	12	0	677	176.8	1,816

[1] All data are median values, except n, which is the number of samples

Table 4.—Results of tests of normality for 57 samples[1]

	Skewness test		Skewness		Chi-square test		
	P<0.05	P>0.05	Left	Right	P<0.05	P>0.05	NA[2]
PRE-REMINING							
Flow rate	45	12	12	45	12	1	44
Acid conc	35	22	22	35	4	7	46
Acid load	42	15	8	49	9	2	46
Iron conc	40	17	11	46	6	4	47
Iron load	45	12	10	47	10	4	43
Sulfate conc	28	24	22	30	4	8	40
Sulfate load	42	10	12	40	11	1	40
POST-REMINING							
Flow rate	41	16	7	50	21	7	29
Acid conc	35	22	13	43	21	6	30
Acid load	40	17	3	54	24	5	28
Iron conc	52	5	6	51	17	8	31
Iron load	42	15	5	52	25	4	28
Sulfate conc	27	27	15	37	12	12	28
Sulfate load	35	17	2	50	15	8	29

[1] A P value greater than 0.05 indicates the data were not normally distributed at the 5% significance level; a P value less than 0.05 indicates that the assumption of normality cannot be rejected with greater than 95% confidence.
[2] Not available. Insufficient degrees of freedom to adequately conduct the chi-square test.

Skewness Test

The skewness test determines normality by comparison of the absolute value of the skewness coefficient for each data set to tables of calculated values based on the appropriate sample size and the significance level (Harris and others, 1987). Positive values of the coefficient indicate the data are right skewed, and negative values, left skewed. For a given sample size, if the skewness coefficient is greater than the tabulated value, the data set is determined to have a nonnormal distribution. Harris and others (1987) tabulated skewness coefficient values at the one-tailed 5% and 1% significance levels on the basis of Monte Carlo random-number simulations using as many as 10,000 randomly generated data sets. Each data set ranged in size from 9 to 30. Tables presented by Snedecor and Cochran (1971) list skewness coefficient values at the 5% and 1% significance levels for data set sizes ranging from 25 to 500. Both table sets were utilized for the skewness test for normality on the data evaluated herein.

Analysis of the skewness test results indicates that the flow rate, concentration, and loading rate data are generally nonnormally distributed at the 5% significance level. A possible exception to this trend is the sulfate concentration. The nonnormally distributed sulfate concentration data only slightly outnumber those that are normally distributed for both pre- and post-remining periods. However, for other parameters, nonnormally distributed data exceed the normally distributed data sets by at least a two-to-one margin. In total, the distribution of pre-remining variables is 277 nonnormal compared to 112 normal. The margin of the post-remining total is slightly less, 272 to 117. The dominating influence of flow on the loading rate is indicated by the differences between the pre- and post-remining distribution of sulfate concentrations and sulfate loads. Sulfate loads are similar to the corresponding flow and dissimilar to the distribution for the sulfate concentrations for each period. This indicates that the flow influence on load overrides the sulfate concentration influence. This strong influence of flow on load determination is extremely important information for the formulation and implementation of abatement techniques intended to reduce the contaminant load.

Table 4 illustrates that the data sets are predominantly skewed right (toward lower values). The ratio of right to left skewed data ranges from a low of approximately 1.4 to 1 for pre-remining sulfate concentration to a high of 25 to 1 for post-remining sulfate load. In general, the number of concentration data sets skewed right for the post-remining period exceeds that for the baseline period. This may be related to the state of flux of the post-remining spoil aquifer, which yields periodic extreme concentration and/or flow values. The skewness form of the loading data sets is more predominantly to the right than

either the flow rate or the concentration data sets. This increase in skewness appears to be caused by the interdependence of concentration and flow (e.g., concentration increases caused by flushing or concentration decreases caused by dilution). The overwhelming majority of the pre- and post-remining data sets tend to be skewed right, 629 out of 777 (81%). These trends are similar to those observed by Montgomery and others (1987) for ground water quality not related to coal mining activities.

Chi-Square Test

The chi-square "goodness-of-fit" test procedure calculates a statistic that permits a comparison of the observed frequencies to those of a known distribution (i.e., normal). As the difference between the observed frequencies and a normal distribution increases, the chi-square statistic becomes larger. The greater the chi-square values, the greater the probability that the two distributions are dissimilar. Some of the data sets consist of relatively few observations, and therefore have insufficient degrees of freedom to adequately conduct the chi-square test. This is illustrated in table 4.

Table 4 indicates that the chi-square test produces similar results overall as the skewness tests. However, because of the nonapplicability of the chi-square test for many of the data sets, a direct comparison is not possible. The chi-square test results illustrate that the majority of the flow, concentration, and loading rate data are nonnormally distributed at the 5% significance level. However, there are some inconsistencies within the contaminant concentration data sets before and after remining.

Chi-square testing of the pre-remining contaminant concentration data sets indicates that they are more commonly normally distributed than nonnormally distributed by a margin of 19 to 14. Conversely, the postremining data tend to be more often nonnormally distributed than the pre-remining data (50 to 26). The pre-remining acidity and sulfate concentration data sets are normally distributed by a two-to-one margin over nonnormally distributed data. The pre-remining iron concentration data exhibit a tendency to be nonnormally distributed at the 5% significance level. The post-remining data are mainly nonnormal for acidity and iron concentrations at the 5% significance level. Post-remining sulfate concentration is equally likely to be normally or nonnormally distributed. The differences in distribution between pre- and post-remining concentration data may be related to the limited number of tests that could be adequately conducted on these data sets compared with the skewness tests.

The chi-square testing indicates that the contaminant loads exhibit similar distributions as the corresponding flow data

flow data (table 4). This suggests that flow rate strongly influences loading rates, as was observed by Smith (1988) for acidity. This is especially evident for the pre-remining acidity and sulfate, for which concentrations are mainly normally distributed. Chi-square tests on the corresponding loading data sets indicate that they are primarily nonnormal at the 5% level, mirroring the trends of the flow data sets. If surface recharge and ground water flow into and through these minesites can be controlled, the mine operator may be able to engineer a reduction in discharge outflow and, in turn, contaminant load during mining or reclamation. Recharge-limiting abatement practices, such as regrading spoil piles to promote surface runoff and sealing of the highwall and exposed mine entries to reduce lateral inflow, should be the most successful in reducing the contaminant load.

EXPLORATORY DATA ANALYSIS

One of the tools of exploratory data analysis is the "notched box-and-whisker" plot. This type of plot is used to graphically display several basic statistical parameters. These plots are useful for a visual comparison of subsets of data (see figures 7 through 10). The bottom and top ends of the box correspond to the first and third quartiles of the data—the central 50% of the data from the 25th to 75th percentiles (interquartile range) are contained within the box. The width of the box is directly proportional to the square root of the number of observations in the represented group. The vertical lines, or "whiskers" on the top and bottom of the box extend to the largest or smallest

observation within 1.5 times the interquartile range. Extreme values lying between 1.5 and 3.0 times the box length beyond the box ends are possible outliers and are represented by squares. Any value greater than 3.0 times the interquartile range from the end of the box is considered an extreme outlier and is

Figure 8
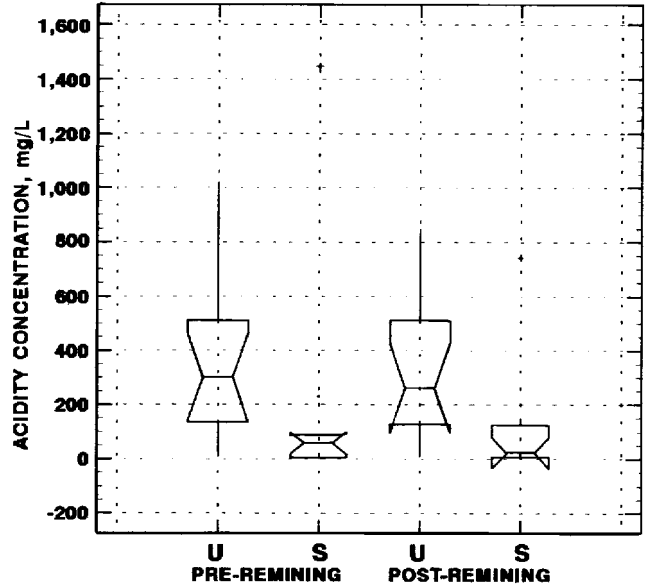
Acidity concentration medians of underground (U) and Surface (S) mines before and after remining.

Figure 7
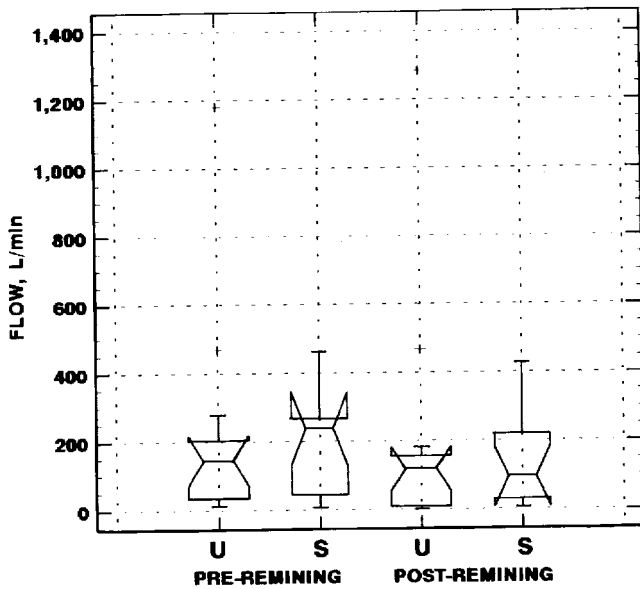
Median flow rate measurements of underground (U) and surface (S) mines before and after remining.

Figure 9
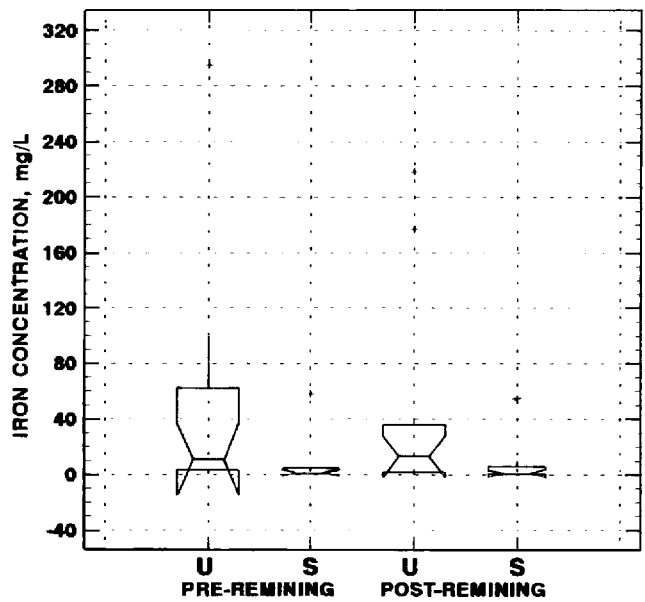
Iron concentration medians of undergound (U) and Surface (S) mines before and after remining.

marked with a plus sign (+). Within each box, the horizontal line denotes the median of the data and notches (indentations) in the sides of the box approximate the 95% confidence interval about the median. The median rather than the mean is evaluated because it is a more consistent indicator of central tendency of nonnormally distributed data. When notched box-and-whisker plots are compared, if the notches about the median of comparable box plots do not overlap, then the medians are said to be significantly different at the 95% confidence level. For additional information on notched box-and-whisker plots, the reader is directed to McGill and others (1978).

Figure 7 is a notched box-and-whisker plot representing the sum of the median flow rate measurements for each site, classified by mine type (underground and surface mines) before and after remining, respectively. The comparison of flow rate characteristics for surface and underground mine discharges before and after remining indicates that there is no significant difference (at the 95% confidence level) of the median values.

Figure 8 exhibits the average acidity concentration determined for each mine using the individual discharge median values. If the site had only one discharge, then the average concentration was equal to the site median. As with figure 7, these data were plotted on the basis of pre-remining and post-remining, underground and surface mine discharges. Figure 8 illustrates that the pre- and post-remining medians of the acidity concentrations for underground mines are significantly higher at the 95% confidence level than the acidity concentrations for

Figure 10

Sulfate concentration medians of underground (U) and surface (S) mines before and after remining.

surface mines. Higher acidity values are exhibited by the underground mines because ground water flows almost exclusively through the portions of the mine where pyritic material (relatively high-sulfur coal, seat, and roof rock) is exposed and AMD forms. Conversely, Hawkins and Aljoe (1991) observed that in surface mines, ground water flows through relatively discrete paths in the highly fragmented poorly sorted spoil material. Portions of the spoil may consist mainly of acid-forming materials (e.g., high-sulfur carbonaceous black shales, sandstones, and spoiled coal), while other parts may consist mainly of alkaline-forming materials (e.g., limestones and carbonate-rich shales).

The underground mine discharges generally exhibit a broader range of values, excluding the outliers and far outliers, than the surface mine discharges in figure 8. The broad range of acidity concentration exhibited by the underground mines may be related to the broad range of site ages. The abandoned underground mines may be up to, and in some few cases more than, 70 years old. In the older sites, natural amelioration of the AMD-forming mechanisms over time may result in lower acidity values as the exposed pyritic minerals are exhausted. Relatively newer underground mines, in the same coal seams and in the same region, may yet be yielding elevated acidity concentrations. Similar natural amelioration was observed by O'Steen and Rauch (1983) at surface mines in northern West Virginia. Natural amelioration processes have had less time to influence the relative severity of discharges from abandoned surface mines because most surface mines date from the early 1970's and so are about 20 years old or less. In figure 8, post-remining acidity concentration is nearly identical to the pre-remining plot, indicating little change occurred relative to the abandoned mine type.

Figure 9 is a plot of the average iron concentration determined from the individual discharge median values. The configuration is similar to that for the acidity concentrations (figure 8), although more subdued. The underground mine iron concentration median is higher than the median for surface mines, although the differences between the medians are not significant at the 95% confidence level. The trends observed for iron concentrations are related to the same causal factors as the trends for acidity. The post-remining iron concentration plot is not significantly different from the pre-remining plot with regard to the mine type.

Figure 10 represents a plot of the average of the pre- and post-remining discharge sulfate concentration medians for each site. As exhibited by the plots of acidity and iron concentrations, the underground mine median values are higher than those for surface mines. The difference between the medians is not significant at the 95% confidence level for the pre-remining data. The pre-remining interquartile ranges are likewise larger for the underground mines. The post-remining
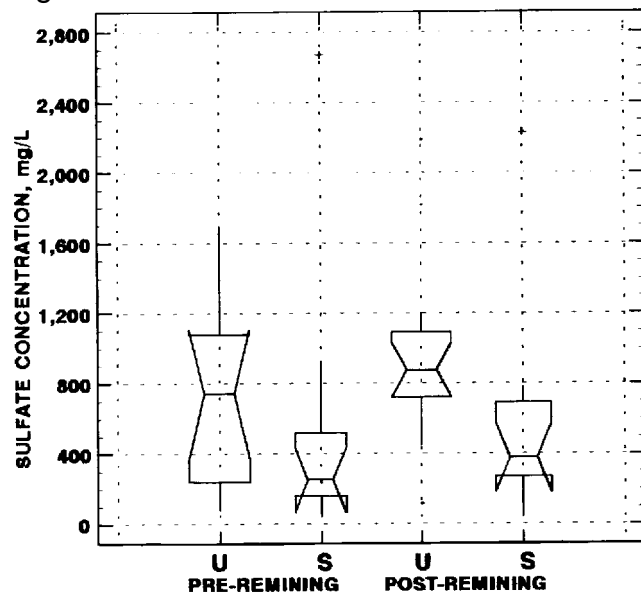
sulfate values, although somewhat similar to the pre-remining values, exhibit a significant difference by mine type at the 95% confidence level. This is caused by a rise in the underground mine median and a narrowing of the approximate 95% confidence interval about the median (notches) of the underground mine data. Figure 10 illustrates that daylighting of the abandoned underground mines greatly decreases the range of variability in sulfate concentration.

Notched box-and-whisker plots (shown in appendix B) were created for pre- and post-remining of the acidity, iron, and sulfate loads for underground and surface mines. The configuration of these plots is similar to those of the corresponding concentration plots, indicating that concentration does to some extent influence load. None of the loadings differ significantly at the 95% confidence level, between underground and surface mines. The plots for flow pre- and post-remining (figure 7) do not exhibit significant differences in the medians at the 95% confidence level. This indicates that flow rate may have a stronger influence on loading than concentration does, which corresponds to trends exhibited by the normality test results. This is especially evident where the contaminant concentration exhibited significant differences at the 95% confidence level.

NONPARAMETRIC CORRELATION

Correlation coefficient is a measure of the interrelationship between two variables from which the degree of statistical significance can be determined. Correlation coefficient is generally determined using parametric testing procedures (Davis, 1986). However, the flow, concentration, and load data in this study are mainly nonnormally distributed. An attempt to transform these data into an approximate normal distribution using a log transformation was unsuccessful. Therefore, nonparametric methods must be used to determine the correlation coefficient.

Spearman's rank correlation is a nonparametric test that determines the similarity or dissimilarity of two data sets. This procedure uses ranked data sets to calculate the correlation coefficient, instead of using the actual data values (Davis, 1986). The correlation coefficient ranges from +1.0 to -1.0, which indicate a perfect positive and negative relationship, respectively. A table of critical values of Spearman's rank correlation (r) is substituted for the standard t-distribution table, because the t-test is based on the assumption that the data are from a bivariate normally distributed population. The table of critical values is used to determine the significance of the correlation coefficient (Davis, 1986). For the purposes of this study, the significance was established at the $P = 0.05$ level.

Spearman's rank correlation coefficient values were calculated to determine the interrelationship that flow and concentration have with the corresponding loading rate. The results, summarized in table 5, once again illustrate that flow is much more often strongly correlated to the contaminant load than is contaminant concentration.

Table 5.—Significant correlations using Spearman's rank correlation for 57 samples[1]

	Pre-remining			Post-remining		
	Acidity	Iron	Sulfate	Acidity	Iron	Sulfate
Flow (+) . . . . versus	42	29	37	51	42	51
Load (-)	0	0	0	0	0	0
Conc (+). . . . versus	10	14	8	30	30	12
Load (-)	5	5	4	7	2	6

[1]Values are number of data sets exhibiting a significant positive (+) or negative (-) correlation at $P = 0.05$ level.

Approximately 82% of the pre- and post-remining acidity loadings (93 of 114 samples) are significantly correlated to the flow rate, while the concentration exhibits a significant correlation (positive or negative) to load for slightly less than half (52 of 114) of the cases. The flow rate correlations are in all cases positive, indicating that flow increases are accompanied by load increases. Significant correlations of flow to acidity load increased moderately from pre- to post-remining. Pre-remining acid concentration exhibits a positive correlation to contaminant load about one-fourth as often as flow. Post-remining acidity concentration is correlated positively to load three times as often as pre-remining acidity data. This increase may be related to the state of geochemical and hydrologic flux that exists for a few years after reclamation. Pyrite oxidation products, formed while spoil is exposed to atmospheric oxygen during mining, tend to get flushed out in "slugs" of contaminant by recharge events in the period following reclamation. In five cases, pre-remining acidity concentration exhibits a negative correlation to load. This may be caused by dilution from high flow events, as with the type 1 discharge described by Smith (1988). Negative correlations between acidity concentrations and load for the post-remining data are not substantially different from those for pre-remining data.

The weakest correlation of flow rate to contaminant load of the three contaminants analyzed is observed for iron. About 51% (29 of 57) of the pre-remining iron loadings are strongly correlated to the flow rate, compared with about 25% (14 of 57) of the cases in which iron concentration exhibits a significant positive correlation to load. After remining, the number of strong correlations of flow to iron contaminant load increases moderately to 74% (42 of 57). After remining, the instances of positive correlation of iron concentration to contaminant load

double to 53% (30 of 57). Five discharges show a negative correlation of iron concentration to load before remining. The number of negative correlations between iron concentration and load (two) is not substantially different after remining.

Almost 89% of the pre- and post-remining samples for flow rate versus sulfate loadings (88 of 99 samples) are significantly correlated (all positive). A lower total number of possible sulfate correlation values occur because pre-remining sulfate data are not available for 15 discharge points. No negative flow-to-load correlations were noted. For sulfate concentration, as for acidity and iron concentrations, the number of strong correlations of flow rate to contaminant load increases from pre- to post-remining. Of all the contaminants, combined pre- and post-remining samples for sulfate concentration versus load exhibit the lowest number of significantly correlated (positive and negative) data, 30 out of 99. Correlations of sulfate concentration to load, both positive and negative, change very little because of remining.

Overall, significant positive correlations of flow rate versus loading outnumber positive correlations of concentration versus loading by over 2 to 1 (252 to 104). The increase of the post-remining positive correlation of flow rate and concentration to load may be related to the previously discussed state of geochemical and physical flux of mine spoil during this period. The water table is in the process of rebounding (reestablishing), while the spoil is undergoing considerable changes that directly affect the transmissive properties of the aquifer. High recharge events will tend to flush out high levels of contaminants from the freshly exposed and oxidized pyrite, while also yielding higher discharge rates.

All negative correlations are exhibited by contaminant concentration to load. This may be caused by dilution of contaminants from increased flow rates; also, in the cases of acidity and iron, geochemical changes of the ground water may reduce concentrations through chemical reaction. During high flow events the sources and flow paths of the ground water may change, thus facilitating water quality changes. Chemical reactions, brought on by ground water quality changes, can reduce the acidity and iron content, but generally will not affect the sulfate content (at the levels of sulfate observed). Because the number of negative sulfate concentration-to-load correlations is similar to those for iron and acidity, dilution and not geochemical reactions appears to be the main cause of most of these negative correlations.

The Spearman's rank correlations indicate, as did the normality tests and the notched box-and-whisker plots, that for determination of contaminant loading rates, flow rate is the main controlling factor. Concentration is a subordinate factor. Thus, a reduction in contaminant load is almost a certainty, if recharge to the ground water can be diminished through mining and/or reclamation (abatement) practices.

PRE-REMINING SAMPLING ADEQUACY

One of the main objectives of this analysis was to determine to what degree a discharge will, from natural processes, exceed a set of simulated effluent standards (SES) for acidity, iron, and sulfate loads. These SES were established by the USBM using the PADER method for 6-, 9-, and 12-month sampling periods and varying sampling frequencies. Acidity and iron were included in this part of the study because they are mandated effluent parameters under the Pennsylvania remining program. Sulfate was included because, as previously stated, it is a relatively conservative indicator of AMD.

To determine the adequacy of baseline background sampling for mine discharge characterization, the 105 remining permit files were reviewed. From these files, 39 permitted operations with a total of 115 discrete discharges were selected for this portion of the project. The study site locations are shown on figure 11. Site selection criteria were based solely on the site having a baseline sampling period of 18 months or more.

The background data were analyzed using the system currently employed by the PADER, as described in the section "Water Quality Standards for Remining." The data were analyzed to determine the optimum sampling frequency and duration that would most accurately characterize baseline contaminant load. A maximum of 1 year sampling duration for establishing the SES was placed on this study. This is based on the assumption that 1 year is the maximum length of time that the majority of mine operators will find acceptable in terms of increased cost and extended permitting time. Proposed remining policy of some coal-producing States has included baseline sampling beyond 12 months, and there is little question that longer sampling time will better characterize the pollution load. However, sampling for 1 year is sufficient to include both high and low flow discharge conditions.

In Pennsylvania, the mine operator collects a temporally consecutive series of pre-remining discharge water samples, along with discharge flow rate measurements. Effluent standards for the remining contaminant loading rate (e.g., pounds of contaminant per day) are established based on the analyses of these data. This baseline background sampling is crucial for the determination of changes to the discharges caused by remining. Insufficient characterization of discharge contaminant load could falsely indicate increased or decreased

Figure 11
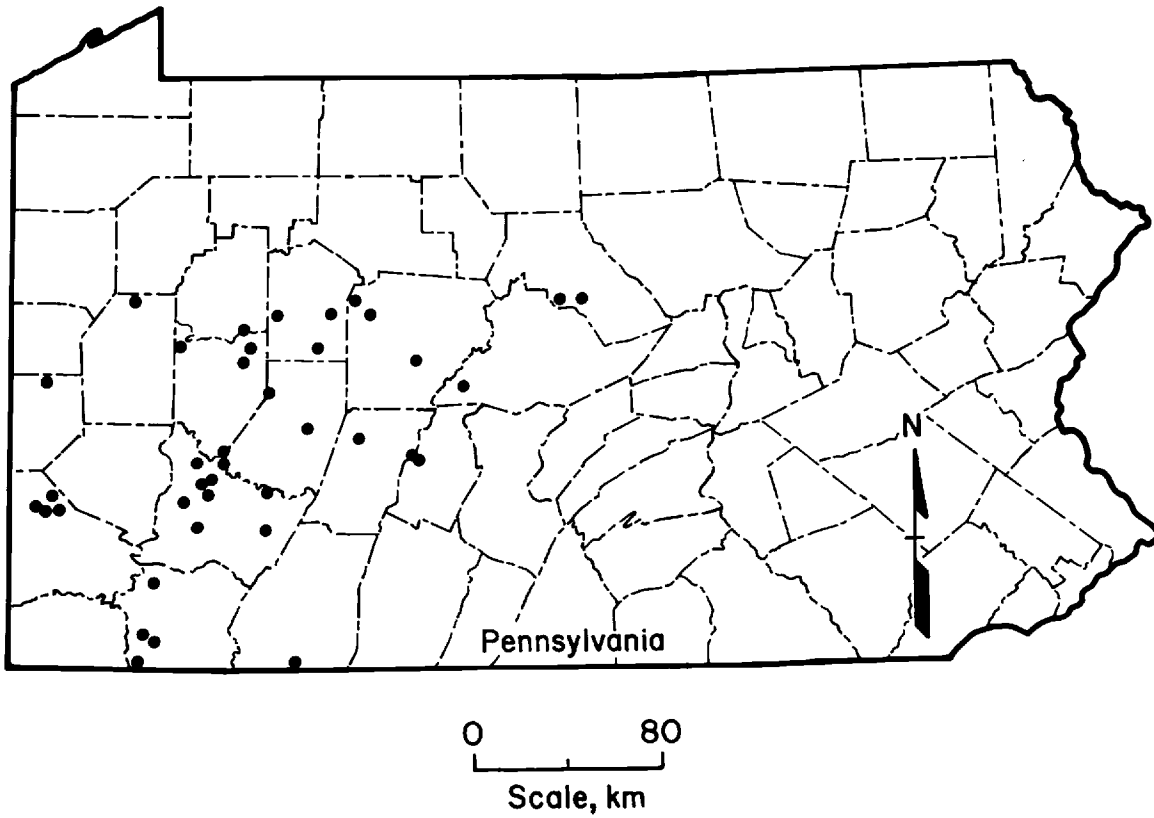
Location of 39 study sites used in study of baseline sampling adequacy.

degradation caused by remining. Subsequently, this false indication could cause a mine operator to be incorrectly held responsible for treatment. Conversely, under characterization could cause an operator to be released from treatment liability when the discharge quality has actually been degraded.

The large number of discharges (115) precluded analysis of individual discharges to determine if the baseline sampling occurred during a period of normal precipitation or an unusually wet or dry period. The lengthy period (18 months or more) over which discharges were monitored should help reduce the impact of abnormally wet and dry periods on the baseline load. It is difficult to ascertain which time interval will accurately represent a discharge in terms of load. However, the sampling represents data collected randomly throughout an 11-year period from September 1980 through November 1991. The sampling of individual discharges ranged from 18 to 86 months within that period. The large number of discharges sampled within the 11-year period may diminish the impact that a protracted wet or dry period or year has on a few discharges. It was not the intent of this portion of the study to determine what sampling period accurately represents individual

discharges, but rather to determine the optimal sampling scenario for remining based on actual data.

DATA ANALYSIS AND DISCUSSION

Pre-remining data from the 115 discharges were used to examine the effect of different sampling frequencies and durations on the baseline contaminant characterization. The number of discharges ranged from 1 to 10 per site, with a median of 2. These 39 mining operations were selected based on the criterion that they possessed 18 months or more of background hydrologic data prior to site activation from mining activities. These operations are primarily remining abandoned surface mines and/or daylighting (surface mining) abandoned underground mines. However, a few of the operations are coal refuse reprocessing operations.

There are a multitude of previous studies on the adequacy of sampling of surface and ground water for situations other than coal mining, which rely primarily on highly complex statistical methodology. Rather than conducting another of these studies, it was decided that an empirical analysis of actual data could

sufficiently determine an effective background sampling scenario to best characterize the discharges. A theoretical statistical study would be highly subjective and would be based on the prevailing thought at that time, whereas analysis of actual data should more objectively analyze the problem and be specific to coal mine discharges.

The sampling frequency varied from less than one to four samples per month. The SES were established using the PADER system of data analyses to characterize the discharges. The study required that each discharge have a minimum of 1 year of monitoring data following the sampling interval used to create the SES and before the site was activated by remining activities. For example, if the SES were calculated based on a 12-month sampling period, at least 24 months of preactivation data were required. The time interval between the SES calculation and site activation was used as a site-specific testing period to determine how often a discharge would naturally initiate treatment under the first two triggering methods of the PADER. The methodology of the second triggering method not only permits determination that the test period data significantly exceed the SES, which triggers treatment, it also permits the determination that the test period data are significantly below the SES. If the test period data are significantly below the SES, this indicates that the SES may have overestimated the contaminant load. Table 6 summarizes the data used to create the SES and the post-SES test period.

Table 6.—Test site and discharge data

SES period	6 months	9 months	12 months
Number of samples used in SES calculations:			
High	13	24	31
Low	2	6	6
Mean	7.5	13	18
Number of samples in post-SES testing period:			
High	114	107	101
Low	6	9	8
Mean	36	40	35
Number of months post-SES samples were taken:			
High	80	77	74
Low	12	12	12
Mean	28	31	27
Number of discharges used in SES calculations	115	81	78

SES   Simulated effluent standards.

These analyses were divided into two parts. The data were first analyzed to determine the optimum sampling duration. One year was considered the maximum acceptable sampling period to the mining industry, because of the extensive permitting time and monitoring expenses. Second, the data were analyzed to ascertain the frequency within the sampling period that would most accurately characterize the contaminant load. The 6-, 9-, and 12- month sampling period data were divided into subgroups based on the number of samples collected in each interval. The optimum duration and frequency portions of this study were analyzed separately using the first two triggering methods of the PADER system.

The first triggering system of the PADER was modified slightly for this part of the study, because the discharges were seldom sampled on a weekly basis. For this study, if a discharge exceeded the approximate 95% tolerance limit for three consecutive months, this was considered a treatment initiation event. It is possible that this modification may overestimate the number of actual treatment-triggering events. However, if a discharge exceeds a 95% tolerance limit for three successive months, this is a valid indication that the background information was inadequate and/or a true change to the discharge has occurred. The intent of this part of the study was to determine the optimal baseline sampling scenario and not to enforce compliance to effluent standards.

The second triggering system was also modified slightly for applicability to this study. The 95% confidence interval about the median was not compared for water years or water-year periods; instead, the 95% confidence interval about the median for the baseline period (6, 9, or 12 months) was compared with the 95% confidence interval about the median for the entire test period, which ranged from 12 to 80 months. However, the analyses must be viewed in the context that there is a potential to narrow the confidence interval about the median as the number of samples increases and vice versa.

DURATION ADEQUACY

The discharge data were analyzed to determine if treatment would be initiated if the SES were the actual permit baseline loading standards and the time interval between the SES establishment and actual site activation (a hypothetical period of remining activities) was the testing period. Additionally, the second PADER triggering method was used to analyze the discharge data to determine if the SES overestimated the baseline contaminant load. In theory, the longer the sampling interval for the SES determination, the more accurate the characterization of the discharge. This is because the longest sampling interval should have the most samples (increasing the statistical validity), and it will include both high and low flow periods in the characterization.

Figure 12, using acidity load, illustrates that with increasing sampling time for the SES, the potential for treatment initiation decreased. At the 6-month sampling period, acidity loads from 37% of the discharges would have initiated treatment using the first triggering method of the PADER. At the 12-month SES, the number triggering treatment decreased to 20%. Similar trends were observed for iron and sulfate loads, as illustrated by table 7.

Table 7.—Optimum sampling duration determination, percent of discharges

SES period	6 months	9 months	12 months
PADER TRIGGERING SYSTEM 1			
Acidity:			
Exceeded SES	37	33	20
Within SES	63	67	80
Iron:			
Exceeded SES	27	22	19
Within SES	73	78	81
Sulfate:			
Exceeded SES	41	35	27
Within SES	59	65	73
PADER TRIGGERING SYSTEM 2			
Acidity:			
Exceeded SES	18	17	18
Within SES	68	75	76
Below SES	14	8	6
Iron:			
Exceeded SES	16	18	17
Within SES	75	78	74
Below SES	9	4	9
Sulfate:			
Exceeded SES	24	31	24
Within SES	65	61	67
Below SES	11	8	9

SES   Simulated effluent standards.

First PADER Triggering Method

The results using the first PADER triggering method indicate that the 12-month pre-remining sampling period best characterized baseline contaminant loads of acidity, iron, and sulfate. This is as expected, based on experience with mine discharge loading rates. This trend appears to be because the 12-month sampling interval included both high and low flow periods.

As the length of the SES period was increased in 3- month increments, the length of the subsequent test period was decreased by an equal amount by default. It is possible that some of the observed decrease in triggering with increased sampling interval length was partially related to this shortening of the test period. However, a review of the data indicated that this effect was minimal. The average length for all of the post-SES test periods was over 2 years (table 6).

Increasing the number of background samples within the SES time period did not significantly decrease the number of discharges initiating treatment, if sampling was not performed on a time-consistent basis. In fact, discharges that had 23 or more samples in 12 months most often triggered treatment. Similar trends were exhibited by the 6- and 9- month sampling periods. The "Frequency Adequacy" section below presents a detailed discussion of this aspect. However, this may to some extent be related to the narrowing of the confidence limits around the median resulting from an increase in the number of samples.

Figure 12
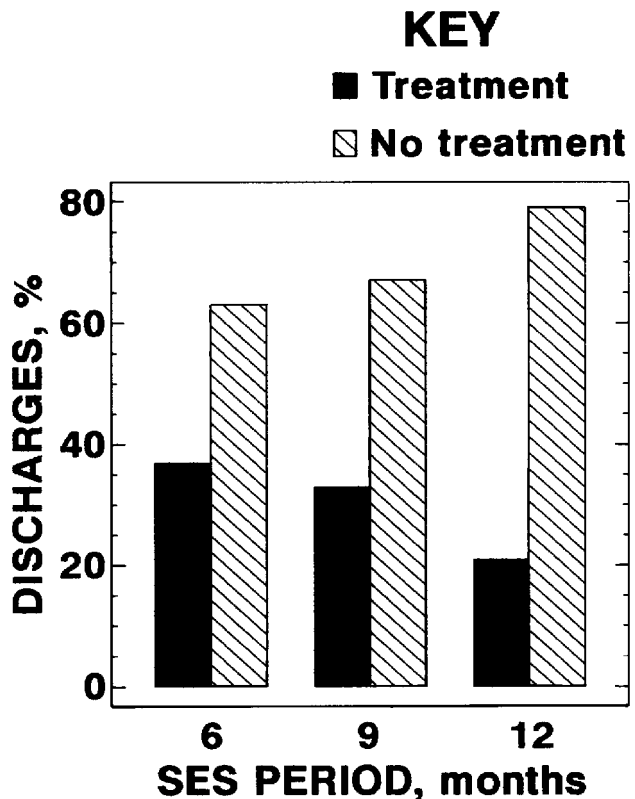
Percentage of discharges initiating treatment for three simulated effluent standard (SES) periods using the first PADER triggering method.

Second PADER Triggering Method

Figure 13 illustrates that the acidity load of approximately a fifth of the discharges (18%) would have initiated treatment at the 6-month sampling interval using the second PADER triggering method. This number changed very little when the sampling interval was increased to 9 (17%) and then to 12 (18%) months. The number of discharges for which iron load triggered treatment with the second method was 16% for the 6-month sampling interval and changed little at the 9- or 12-month sampling level (table 7). Sulfate load triggered treatment nearly 24% of the time for the 6- and 12-month sampling intervals, while the percentage was slightly higher (31%) at the 9-month interval.

With the second triggering method, the number of discharges with contaminant loads that were significantly below the SES was generally smaller than the number exhibiting loads that were above. This is somewhat related to the precipitation and subsequent recharge during the SES period and the test period. However, the large database (115 discharges) and

Figure 13

Percentage of discharges initiating treatment for three simulated effluent standard (SES) periods using the second PADER triggering method.

the lengthy period over which the testing occurred (11 years) should minimize any potential bias of protracted dry or wet periods. The number of discharges that exhibited an acidity load significantly below the SES 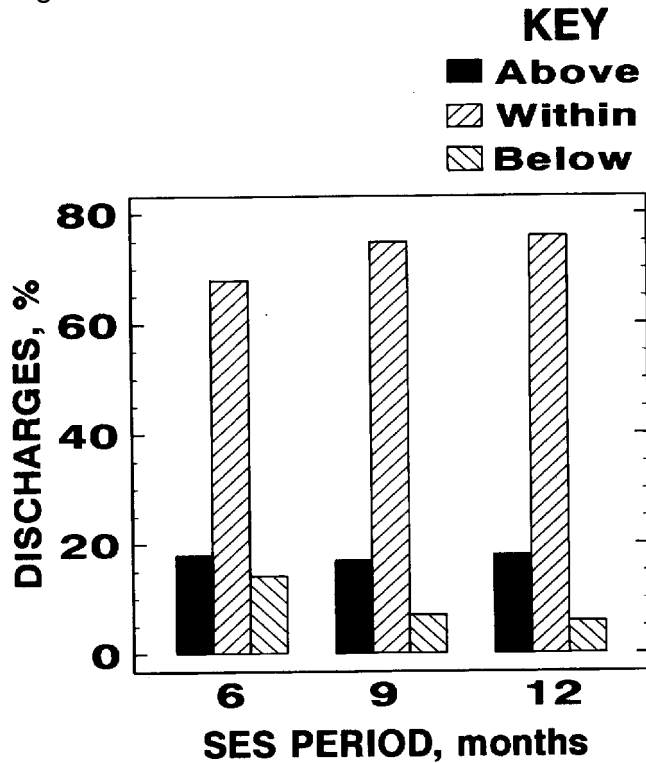decreased as the sampling period increased. Iron and sulfate loads showed no definite trends with increased sampling period length.

With the second triggering method of the PADER, it is somewhat inconclusive as to which of the SES sampling intervals best characterizes the mine discharges. However, the results generally do not contradict the results obtained using the first triggering method. The greatest number of discharges with acidity and sulfate loads within the SES were found at the 12-month sampling level. Iron, on the other hand, changed very little regardless of the sampling interval. The number of discharges exceeding the SES with the second triggering method were in all cases fewer than those found with the first method. Approximately 18% of the acidity loads, 17% of the iron loads, and 24% of the sulfate loads (if sulfate were an effluent parameter) from the discharges would have falsely initiated treatment using the second method.

FREQUENCY ADEQUACY

To determine an adequate sampling frequency, each SES period was divided into four subgroups based on the number of samples collected. The SES periods were divided into nearly equal subgroup sizes. Also, the number of samples collected for one of the four subgroups was equal to the number of months used for the collection period (i.e., an average of one sample per month). As with the duration data, the frequency data were analyzed using the first two PADER triggering methods.

First PADER Triggering Method

The frequency data for the first triggering method are summarized in table 8. The results illustrate that at the 6-month SES level, there was little difference in the initiation of treatment based on acidity or iron loads among the number of samples taken. However, it appears that sulfate (if it were used an effluent standard) would have increased triggering with increasing sample size.

Table 8.—Frequency adequacy data from the first PADER triggering mechanism, percent

Number of samples	6 months			
	<6	6	7-10	>10
Acidity:				
Treatment	38	33	40	37
No treatment	62	67	60	63
Iron:				
Treatment	27	25	23	33
No treatment	73	75	77	67
Sulfate:				
Treatment	27	42	50	48
No treatment	73	58	50	52
	9 months			
	<10	10-12	13-17	>17
Acidity:				
Treatment	43	29	24	38
No treatment	57	71	76	62
Iron:				
Treatment	22	29	12	31
No treatment	78	71	88	69
Sulfate:				
Treatment	30	41	24	50
No treatment	70	59	76	50
	12 months			
	<13	13-17	18-23	>23
Acidity:				
Treatment	25	17	8	40
No treatment	75	83	92	60
Iron:				
Treatment	25	13	12	33
No treatment	75	87	88	67
Sulfate:				
Treatment	23	26	17	47
No treatment	77	74	83	53

At the 9-month SES level, treatment triggering by acidity, iron, and sulfate loads occurred for the lowest number of discharges in the 13- to 17-sample range (24%). The next lowest number of discharges triggering treatment were in the 10 to 12 range for acidity load and the less- than-10 level for iron and sulfate load. However, no definite trends were exhibited at the 9-month SES level as the sample set size changed.

At the 12-month SES level, the 18- to 23-sample size range exhibited the least number of discharges triggering treatment (under 20%) for all three contaminants. However, the triggering levels for the 13 to 17 range and the less-than-13 level were similar, roughly 25% or less. The greater-than-23 level exhibited the highest triggering rate (one-third or more) for all three contaminants, indicating that larger sample set sizes may not more accurately characterize mine discharges. This may be caused by narrowing of the tolerance limits about the median related to an increase in the number of samples. Another explanation is that the greater-than-23 level, in some cases, had an inconsistent sampling frequency, which caused unequal weighting of wet or dry periods.

Second PADER Triggering Method

The results of triggering based on the second PADER method are somewhat different from the results obtained using the first triggering method. The least number of triggering events occurred mainly in the smallest sample set sizes (<6, <10, and <13) for each of the three SES periods (table 9). However, at the 6-month SES, the results for the six-sample size are similar to results for the less-than-six-sample size. The low number of triggering events occurring in the smallest sample sizes could be related to the narrower confidence interval about the median with increasing size of the test period data set.

The results for the number of discharges below the SES levels using the second method are somewhat inconsistent for the three sampling intervals. For the 6-month SES period, the least number of discharges below the SES for iron and sulfate were in the six-sample set size (table 9). However, for acidity, the lowest number of excursions below the SES was for the 7- to 10-sample set size. On the other hand, the 9- and 12-month sampling intervals exhibited the least number of excursions above the SES for acidity, iron, and sulfate loads in the higher sample set sizes for each sample interval.

The results from the second triggering method, although somewhat mixed, appear to indicate that sampling on a consistent monthly basis may be optimal. At the 6-month SES period, the number of excursions above or below the SES for acidity, iron, and sulfate were lowest in the less-than-six and six-sample sizes. The least number of excursions above or below the SES occurred in the smallest sample set sizes for 9- and 12-month SES periods, which are primarily composed of

sample sets where a sample per month was taken. Furthermore, the highest number of discharges within the SES standard (six of nine) occurred when the sampling interval averaged once per month (table 9).

Table 9.—Frequency adequacy data from the second PADER triggering mechanism, percent

Number of samples	6 months			
	<6	6	7-10	>10
Acidity:				
Exceeded SES	3	12	30	30
Within SES	82	71	60	55
Below SES	15	17	10	15
Iron:				
Exceeded SES	6	4	33	19
Within SES	82	96	57	70
Below SES	12	0	10	11
Sulfate:				
Exceeded SES	10	17	40	27
Within SES	73	75	50	65
Below SES	17	8	10	8
	9 months			
	<10	10-12	13-17	>17
Acidity:				
Exceeded SES	8	23	13	31
Within SES	84	65	83	63
Below SES	8	12	4	6
Iron:				
Exceeded SES	13	23	13	31
Within SES	83	71	83	69
Below SES	4	6	4	0
Sulfate:				
Exceeded SES	19	47	29	31
Within SES	72	41	67	63
Below SES	9	12	4	6
	12 months			
	<13	13-17	18-23	>23
Acidity:				
Exceeded SES	13	22	8	33
Within SES	67	69	92	67
Below SES	20	9	0	0
Iron:				
Exceeded SES	20	13	16	20
Within SES	67	74	76	80
Below SES	13	13	8	0
Sulfate:				
Exceeded SES	8	22	28	33
Within SES	75	61	72	60
Below SES	17	17	0	7

SES   Simulated effluent standards.

It is probable that sampling two or four times per month at a consistent time interval will yield similar if not slightly better results. However, there were an insufficient number of discharges where the sampling was two or four times per month to yield conclusive results. Given the good results when sampling occurred once per month, it may not be cost effective

to increase the frequency to gain a slight improvement in discharge characterization.

This triggering method indicates that sampling on a consistent basis (e.g., the third Monday of each month) yields the best characterization of the discharges. This method should avoid overemphasizing either a wet or dry period. This is especially true if the sampling is performed over a complete water year. The results of the first triggering method, although somewhat dissimilar, do not contradict this assertion. In either case, the highest sampling frequency did not yield the best characterization results. This may be because the highest sampling frequency can allow unequal sampling during a wet or dry period, thus biasing the data.

DETERMINATION OF CONTAMINANT LOAD REDUCTION

To determine the effectiveness of remining in terms of reduction in contaminant load, this study used the 57 discrete discharges from 24 western Pennsylvania remining operations that were analyzed in the "Discussion of Statistical Methods" section. These sites are exhibited on figure 6. Table 3 is a summary of the raw flow and concentration data from these 24 sites. The loading data are exhibited in appendix A. As previously stated, these sites were selected because they had been completed (backfilled to rough grade) and possessed a minimum of 1 year of post-backfilling water quality and flow data. The pre- and post-remining data were analyzed using exploratory data analysis (schematic summary), which is the method currently used by the PADER; the Mann-Whitney U test; and a method of nonparametric upper prediction limits (NUPL) (Gibbons, 1990). The test results were secondarily analyzed to assess the applicability of each of these analytical methods to these types of data to determine the effectiveness of remining in contaminant load reduction.

RELATED STUDIES

Crucial to any remining program are the methods of analysis used to determine possible changes in the discharge water quality with respect to the pre-remining conditions. In order to choose an applicable analytical method, characteristics of the water quality and discharge flow must be thoroughly understood. There have been a multitude of studies pertaining to the statistical analysis and characterization of ground and surface waters unrelated to degradation from coal mining. Statistical studies pertaining to coal mine water have primarily been limited to temporal studies of contaminant concentration changes caused by mining within a previously unmined watershed.

Harris and others (1987) reviewed various statistical methods used to characterize ground water quality. They concluded that the skewness test was the most applicable for the determination of normality. They recommended using the Mann-Whitney U, Student's t, ANOVA, and Kruskal-Wallis tests to detect seasonality effects in the data. Simple autocorrelation was suggested to determine serial dependence of data. Montgomery and others (1987) applied these determinations to actual water quality data in a companion paper. They determined that ground water quality data are by-and-large nonnormally distributed and positively skewed (right), often exhibit seasonal fluctuations, and can be serially correlated in time, if sampled on a quarterly basis.

There have been few previous studies pertaining to the impacts of remining on the ground-water quality or quantity, because true remining is a relatively new procedure. Previous remining studies have generally been limited to feasibility or case studies.

Richardson and Dougherty (1976) investigated the technical and economic feasibility of daylighting (surface mining by removal of the overburden) an abandoned underground mine in Garrett County, MD. They collected water quality data before and after daylighting in order to evaluate its impact.[3] They observed that the post-remining contaminant loads were not significantly different from the pre-remining levels, although the post-remining contaminant concentrations exhibited seasonal fluctuations that were significantly higher than those of the pre-remining levels. At the time the report was written, the post-remining water quality data indicated that a slight improvement in contaminant load had occurred compared with pre-remining conditions.

Reed (1980) analyzed the effect of daylighting on a 344-ha (850-acre) underground mine in Tioga County, PA. The report was prepared while the site was being actively mined. Approximately 12% of the mine had been daylighted. He observed that the remining was causing a significant increase in acidity concentration in the discharges draining the affected areas. There appeared to be a direct correlation between the amount of daylighting and the increase in acidity concentration.

Helsel (1983) analyzed streams from mined and unmined watersheds in eastern Ohio to determine the influence of mine and rock type on water quality. He determined that overburden lithology of the mine influenced several water quality constituents. The water quality exhibited neither a normal nor a lognormal distribution. The effects of mine and rock type on water quality were more adequately represented by analysis of the ranked data, rather than analysis of the actual data. Helsel (1987) illustrated the advantages of using nonparametric procedures on water quality data. These advantages include the following: data transformations are not required, the tests can be performed even if normality of data sets is not achievable,

---

[3]Ackerman, J. P., P. S. Campion, and E. B. Persson. Preliminary final rep., Deer Park Daylighting Project. U.S. EPA, undated.

there is greater power in highly skewed data sets, central-tendency comparisons are made on the median rather than the mean, and below-detection-limit data can be easily incorporated without bias. The potential bias is reduced by the use of the data median, which is less sensitive than the mean to a few extreme outlying values (Snedecor and Cochran, 1971).

O'Steen and Rauch (1983) analyzed the spatial and temporal ground water quality variability associated with surface mining in West Virginia. Peak ground water contamination in terms of sulfate occurred approximately 3 years after mining was initiated. Sulfate contamination declined slowly after the peak was reached and was still significant in shallow ground water after 20 years. Razem (1983), to a lesser extent, studied the temporal and spatial effects on ground water of surface mining in a small watershed in Ohio. The ground water in the spoil zone was observed to be "significantly poorer" after mining. Lindorff (1980) studied the long-term effects of surface mining on ground water at an Illinois coal mine. He concluded that, even after approximately 40 years, the ground water quality was "more mineralized than one would expect for undisturbed overburden."

DATA ANALYSES AND RESULTS

In general, hydrologic data have been shown to be asymmetric and nonnormally distributed. Therefore, the most applicable methods of analyses are nonparametric statistics (Hirsch and Slack, 1984; Montgomery and others, 1987; Harris and others, 1987; Berryman and others, 1988). Helsel (1983) likewise showed that mine discharge hydrologic data, such as drainage quality, discharge flow rate, and loading rates, are generally not normally distributed and are commonly positively skewed toward the lower values.

First PADER Triggering Method

Table 10 summarizes the number of sites that triggered the PADER system for weekly sampling (first triggering method) and subsequent treatment for acidity, iron, and sulfate. Two of the sites did not have data for pre-remining sulfate concentration; therefore, sulfate "standards" were established for only 22 of the sites. For this portion of the analyses, using the PADER system, discharges from each site were combined to evaluate overall load changes on a site-by-site basis.

However, for the Mann-Whitney U test and the NUPL, the discharges were analyzed separately in an attempt to provide a more detailed assessment of the impacts of remining.

Table 10.—Weekly sampling frequency and discharge treatment triggering after remining

	Triggered weekly sampling	Triggered treatment
Acidity:		
Yes	10	3
No	14	21
Iron:		
Yes	5	3
No	19	21
Sulfate:		
Yes	11	5
No	11	17

NOTE.—"Yes" indicates that triggering occurred at least once to one or more discharges or hydrologic units for that site. "No" indicates triggering never occurred.

Acidity levels initiated weekly water sampling at least once during the post-remining period for 10 of 24 sites. Three of these sites subsequently triggered treatment. Iron was the triggering contaminant for weekly sampling in 5 of 24 sites. Of these five sites, three secondarily triggered discharge treatment. Of the three incidences of discharge treatment for acidity and iron, two occurred concurrently; therefore, treatment was actually initiated at least once on four separate sites. The elevated contaminant loads that triggered treatment were in all cases transient events. The higher loads generally occurred shortly after reclamation (less than 1 year), and the levels usually declined to within standards within a brief period (less than 6 months). Sites 4, 8, and 19 (table 3) were the mines that had at least one discharge that triggered treatment for acidity on one or more occasions. Mine 8 was the only site to trigger treatment 13 months or more after reclamation. There were no incidences on these sites that required treatment for longer than 6 months.

Sulfate would have triggered weekly sampling, if it were a regulated effluent contaminant, for half the sites (11 of 22). Almost half (5 of 11) of the weekly sampling events would have subsequently initiated treatment. The greater number of sites where sulfate loads, rather than the acidity or iron, would have (if regulated) initiated weekly sampling and, subsequently, treatment may be related to the extreme hydrologic and

geochemical changes that occur in surface mine spoil immediately following mining and reclamation. The rock surface area, hence pyritic material, exposed to oxidation is greatly increased by the mining and reclamation processes. This promotes acid production, which is indicated by elevated sulfate levels. Concurrently, the increased rock surface area also increases the exposure of alkaline strata (e.g., limestones, dolostones, and calcareous shales), when present, to weathering and dissolution, thus adding alkalinity to the ground water system. The added alkalinity reduces the acidity concentration and raises the pH. As pH levels rise, the potential for dissolved iron to oxidize and precipitate out of solution increases. Sulfate concentrations are little affected by increases in alkalinity and pH at the sulfate and calcium levels common to mine water. Therefore, it is possible for mine water to exhibit significant increases in sulfate without corresponding increases in acidity or iron.

The results shown in table 10, created using the first triggering method of the PADER system, illustrate that discharge treatment was seldom incurred (a total 4 of 24 sites for combined acidity and iron). A review of the cases where treatment was initiated indicate that the treatment was of an ephemeral nature and occurred most often shortly after reclamation.

Second PADER Triggering Method

The second method used by the PADER to determine if the remining has further degraded the mine water compares the post-remining median load of a water year with the median for the baseline data. Lack of overlap at the 95% confidence interval for the pre- and post-remining median indicates that the medians are significantly different at the 5% significance level. The rationale for use of a water year is to ensure that the data used in the comparisons are not biased by the occurrence of most of the sampling during a particularly low or high flow period. The large number of possible water years and water-year periods for 24 sites with 57 discharges precluded the strict adherence to this part of the PADER system for this study. Instead, the median of the data for the entire post-remining period was compared with the median of the pre-remining data. Any possible sampling bias of the data is minimized because each of the post-remining data sets includes from 2 to over 10 of both low and high flow periods. The use of the median rather than the mean further minimizes possible bias (Snedecor and Cochran, 1971).

Results from using this method indicate that for a majority of these sites, remining did not cause additional degradation of the mine discharges (table 11). The median acidity and iron loads for 21 of 24 sites was equal to or below the baseline levels at the 5% significance level. The one site where acidity was above pre-remining levels coincided with one of the sites

of the iron excursions. Therefore, there were a total of three sites that indicated possible degradation caused by remining. This illustrates that over 87% of the sites exhibited no significant degradation in terms of acidity and iron. Median load declined to less than the pre-remining median for 7 of 24 sites for acidity and for 4 of 24 sites for iron. This indicates that nearly a third of the sites had a statistically significant improvement in acidity load. For three of these sites, both acidity and iron were significantly below the baseline levels, making a total of eight sites with an improvement in the effluent load. Of the 11 excursions where acidity and iron loads were significantly below the pre-remining median, 7 sites likewise exhibited a significant flow reduction. However, concentration appeared to have contributed in six of these excursions. The median concentration dropped by a factor of 2 or more for these six excursions.

Table 11.—Comparison of post-remining minesite median contaminant loads to pre-remining median at the 5% significance level

Load	Above	Within	Below
Acidity	1	16	7
Iron	3	17	4
Sulfate	3	15	4

Nearly 13% (3 of 24) of the sites exhibited median acidity and iron loads above the pre-remining median, indicating an apparent significant increase in contaminant load. Of the four instances where the post-remining median contaminant load (acidity and iron) exceeded the pre-remining median, none exhibited a significant increase in flow rate at the 5% significance level. This indicates that when a significant increase in load occurred, it was not related to flow alone; concentration also played an important role. Three of the four excursions exhibited a substantial increase (a factor of 5 or greater) in the concentration median.

The median sulfate load comparison exhibited similar results as acidity and iron load comparisons. Flow rate changes were a significant factor for three of the four excursions below the 5% significance level. However, none exhibited a decrease in the median concentration over 11%. For the excursions above the 5% significance level, flow was never significantly increased, while two of the three exhibited a substantial increase (a factor of 6 or greater increase) in concentration.

Discharge flow rate plays a critical role in loading rate excursions outside of the 5% significance level, especially for decreases in contaminant load. Apparent changes in flow may in some cases be caused by inadequate baseline sampling. If the pre-remining sampling period was unusually wet or dry, this will also create a bias in the data and can incorrectly cause the post-remining data to exhibit an apparent decrease or increase in contaminant load, respectively. This aspect of

sampling is discussed in detail in the "Pre-Remining Sampling Adequacy" section. However, if the pre-remining sample duration and frequency are adequate, the potential error is greatly reduced. Because of the dominant role of flow in determining contaminant load, recharge-limiting abatement practices should be the most successful in reducing the load.

The analyses indicate that concentration is commonly less important than flow rate for the determination of contaminant loading excursions outside of the 5% significance level. However, for all contaminants, substantial changes in concentration level are more often associated with excursions above (71%) rather than excursions below (40%) the 5% significance level.

MANN-WHITNEY U TEST

The Mann-Whitney U (MW) test is a nonparametric substitute for the Student's t-test used to determine if two samples (data sets) have equal means (Davis, 1986). The MW test is an unpaired test in which the test statistic is based on the sum of the ranks of the combined data sets. The MW test should be employed when the data sets exhibit a strongly nonnormal distribution and are of different sizes (Harris and others, 1987). For this study, the MW test was used to determine if the medians, rather than the means, of the data sets were significantly different. The median is not as sensitive to data extremes as the mean and is therefore more commonly used to represent the central tendency of strongly skewed data (Snedecor and Cochran, 1971).

To ascertain possible changes in the mine discharges caused by remining, the MW test was used to compare the pre-remining and the post-remining data sets. This test was conducted separately on the discharge flow rate, contaminant concentration, and loading data. The MW test was used to determine if the medians of the two data sets come from different populations at the 5% significance level. The MW test was applied to individual discharges, rather than on a minesite basis, in an effort to provide a clearer determination of the effect of remining on the water quality. Mine discharge quality and flow can vary widely within a site; therefore, analyzing individual discharges should promote a more accurate assessment of hydrologic changes caused by remining.

The MW test results (table 12) indicate that concentration and load of acidity and iron were unchanged or decreased (improved) at the 5% significance level for the majority (84% to 93%) of the discharges. These results suggest that, in terms of acidity and iron concentration and load, remining generally does not degrade the mine discharge waters. Approximately 30% of the discharges exhibited a significant improvement in

terms of acidity and iron load. The number of discharges above the 5% significance level (indicating degradation) is lower for acidity and iron loads than for the corresponding concentrations. This appears to be related to the strong influence that the discharge flow rate has on the contaminant load, as discussed earlier. The MW test results for acidity and iron loading are very similar to those exhibited by the flow rate and dissimilar to those for the corresponding concentrations.

The number of discharges exhibiting increased sulfate concentration and load from remining is substantially higher than the number of discharges exhibiting increased acidity or iron values. Over a third of the discharges (35%) have an increase in sulfate concentration, and almost 20% have an increased sulfate load at the 5% significance level. This higher "failure" rate exhibited by sulfate is most likely due to the changes in ground water flow paths and contacted material caused by remining and reclamation, as previously discussed. This does not necessarily imply that the remining has increased the contaminant problem, because sulfate is not a regulated effluent parameter and acidity and iron do not show similar trends. The great range of values exhibited by sulfate concentration may be the reason that flow rate is somewhat less dominant in the determination of sulfate loading rate changes than in the determination of changes in acidity or iron.

Table 12.—Mann-Whitney U test comparing post-remining with pre-remining data

	Below[1]	Above[2]	Within[3]
Flow rate	16	4	37
Acid conc	24	7	26
Acid load	17	4	36
Iron conc	18	9	30
Iron load	17	5	35
Sulfate conc . .	9	18	25
Sulfate load . . .	14	10	28

[1]Median was below the corresponding pre-remining median at the 5% significance level.

[2]Median was above the corresponding pre-remining median at the 5% significance level.

[3]Median did not exceed the corresponding pre-remining median at either the upper or lower significance level.

NONPARAMETRIC UPPER PREDICTION LIMITS

A method of NUPL was developed by Gibbons (1990) to detect degradation of ground water caused by waste disposal facilities. The NUPL method is based on the multivariate hypergeometric distribution function. NUPL determine the probability that at least one of the next specified number of contaminant concentration measurements (resamples) will

be less than the maximum contaminant concentration of a background sample set. The probability determination is not dependent on the order of the results, but does require that the samples be independent. As the number of monitoring points and resamplings increases, the number of terms in the probability sum becomes very high, which makes calculation extremely cumbersome. The more easily derived Bonferonni inequality is substituted because it provides an excellent approximation of these probabilities while avoiding the high number of terms (Gibbons, 1990). The option of resampling is crucial to this method. With the use of resampling, a 5% significance level can be obtained with a reasonable number of background samples (in most cases between three and five). This method was originally developed to determine degradation of ground water taken from several monitoring wells, before and after waste disposal or above and below the disposal site. However, the methodology and underlying assumptions permit its use on the hydrologic data of individual discharges of remining sites.

The remining loading data were analyzed by the use of tables and, as required, the equation derived by Gibbons. The upper limit probability was established at the 5% and 1% significance levels (95% and 99% confidence, respectively). Comparisons were made of the contaminant load and discharge flow rate of each discharge for the pre-remining (background) and the post-remining (resampling) data. Individual discharge points were analyzed separately because, as previously stated, significant differences of water quality and especially flow rate can exist between discharges within a minesite.

To achieve 5% and 1% significance levels with 2 post-remining resamplings and 1 discharge point, 5 and 12 respective background samples are required. Raising the number of resamplings to three lowers the number of background samples required to reach these significance levels to three and seven, respectively. Two of the sites lacked sufficient background samples to achieve the 5% level with two resamplings. The number of sites lacking sufficient background data rose to eight at the 1% significance level with two resamplings. With three resamplings, the number of sites with insufficient background samples was reduced to one and two for the 5% and 1% significance levels, respectively.

The contaminant load levels for most of the discharges were below the 5% and 1% significance levels for both the two- and three-resampling scenarios (table 13). The number of discharges below the 5% significance level increased slightly, from 81% to 87%, when the number of resamplings was raised from two to three. However, at the 1% significance level, there was no change. The increase is caused in part by the additional number of discharges (small background sample set sizes) that are able to be analyzed with three resamplings. The increased

number of discharges below the confidence levels with three resamplings is also caused by several discharges that had two successive samples exceeding the background maximum, but failed to have three at the 5% significance level.

Table 13.—Nonparametric upper prediction limits on the loading data

Significance level	Did not exceed		Did exceed	
	5%	1%	5%	1%
Discharge with 2 samples:[1]				
Flow	38	26	9	5
Acidity	39	26	8	5
Iron	42	28	5	3
Sulfate	29	21	13	5
Discharge with 3 samples:[2]				
Flow	49	44	3	3
Acidity	47	42	5	5
Iron	47	42	5	5
Sulfate	37	32	10	10

[1]Two successive post-remining samples exceeding pre-remining maximum.
[2]Three successive post-remining samples exceeding pre-remining maximum.

Of the three contaminant loadings, sulfate most often exceeded the 5% and 1% significance levels (table 13). This is the same general trend exhibited when the sulfate data were analyzed using the PADER system and the Mann-Whitney U tests. The reasons for the higher rate of failure in terms of sulfate were previously discussed.

The trends of the flow rate are very similar to those of acidity and iron loads, indicating that flow rate had a dominating influence on the contaminant loads, as was also observed by Smith (1988). Flow rate is also a strong influence on the sulfate loads, but extreme changes in sulfate concentration levels caused by remining are of a sufficient magnitude that the flow influence is diminished compared with the influence exhibited by acidity and iron loads.

In total, eight sites exceeded the 5% significance level for acidity and/or iron at least once after remining. Five of those sites also exceeded at the 1% level with two resamplings. With three resamplings, five sites exceeded the 5% and 1% significance level for acidity and/or iron.

COMPARISON OF THE METHODS OF ANALYSIS

A site-by-site comparison of the three analytical methods (table 14) for acidity loads yielded generally similar results. Similar results were likewise exhibited in a comparison for iron and sulfate (shown in appendix C). An indication of

degradation under one of the two PADER methods was generally reflected in the Mann-Whitney U test and/or the NUPL method. However, there were a few sites where degradation was indicated by the Mann-Whitney U test or the NUPL method but not by the PADER methods. This result is partially caused by differences in the methods of data analysis and arrangement. Under the PADER system, each site was analyzed as a single hydrologic unit. For the Mann-Whitney U test and the NUPL method, each of the discharges was analyzed separately. An indication of degradation, as shown in table 14, using the Mann-Whitney U test or the NUPL method indicates that one or more discharges from that site exceed the applicable standard. On some sites (4, 14, and 16), one discharge indicated increased acidity, while the rest were within expected limits or decreased. For this reason, acidity load for the corresponding site may show no change or a net decrease using the PADER system.

Minesites with discharges exhibiting both decreases and increases in contaminant load were not unexpected because with remining the ground-water flow paths are commonly altered, thereby causing discharge flow rates to change dramatically and/or discharges to relocate. This is one reason for differences between results from these two analyses (Mann-Whitney U and NUPL) and the two PADER system methods, which look at the site as a single hydrologic unit.

The comparison of analytical methods suggest that the Mann-Whitney U test and the NUPL methods are as applicable to these data as the system presently employed by the PADER. Using a triggering mechanism and framework similar to those of the PADER system, either of these methods should adequately determine degradation or nondegradation due to remining.

Table 14.—Analyses of acidity load before and after remining

Site	PADER system type		Mann-Whitney U test	Gibbons NUPL method			
	1	2		2 samples		3 samples	
				5%	1%	5%	1%
1 ...	N	N	Y	N	N	N	N
2 ...	N	N	N	Y	Y	N	N
3 ...	N	N	N	N	N	N	N
4 ...	Y	N	Y	Y	X	Y	Y
5 ...	N	N	N	N	N	N	N
6 ...	N	N	N	N	N	N	N
7 ...	N	N	N	Y	Y	N	N
8 ...	Y	N	Y	Y	X	Y	Y
9 ...	N	N	N	X	X	N	X
10 ..	N	N	N	N	N	N	N
11 ..	N	N	Y	N	X	N	X
12 ..	N	N	N	Y	Y	N	N
13 ..	N	N	N	Y	Y	Y	Y
14 ..	N	N	Y	N	N	N	N
15 ..	N	N	N	N	X	N	N
16 ..	N	N	Y	X	X	X	X
17 ..	N	N	N	N	N	N	N
18 ..	N	N	N	N	N	N	N
19 ..	Y	Y	Y	Y	Y	Y	Y
20 ..	N	N	N	N	X	N	N
21 ..	N	N	N	N	N	N	N
22 ..	N	N	N	N	N	N	N
23 ..	N	N	Y	N	X	N	N
24 ..	N	N	N	N	N	N	N

NUPL   Nonparametric   upper   prediction   limits.

PADER   Pennsylvania Department of Environmental Resources.

NOTE.—N (no) indicates treatment was not initiated under the PADER system type 1, the post-remining median did not exceed the pre-remining 95% confidence interval about the median under the PADER system type 2, the Mann-Whitney U test indicated no significant difference at the 5% confidence level, and the Gibbons NUPL method did not have two or three consecutive samples exceeding the pre-remining maximum. Y (yes) indicates that treatment was initiated under the given system. X indicates data were not available or were insufficient to complete the analyses.

SUMMARY AND CONCLUSIONS

Remining Program

The remining program in Pennsylvania has been operational for approximately 10 years. Throughout this period, the amount and type of information required for permitting have been modified in order to more accurately reflect the site history, hydrogeologic conditions, and the proposed remining operation. Over 100 remining permits have been issued since 1983.

Remining in Pennsylvania is widespread in the bituminous coal fields and is occurring on virtually all minable coal seams. However, the bulk of the remining (about 76%) is on the Kittanning, Freeport, and Pittsburgh Coal seams. Many of the sites (45%) include multiple-seam mining. Several coal refuse reprocessing operations (10%) have also been issued permits under the remining program.

Mine size ranges widely in total area, abandoned area, and abandoned area to be reclaimed. Nearly 70% of the abandoned area within the permit boundary is reclaimed during remining. The average area of abandoned surface mines to be reclaimed is almost 80% higher, on a per-site basis, than the area of abandoned underground mines to be reclaimed. This is related to the large quantities of ground water to be encountered, the lower total coal yield, and higher uncertainty associated with remining underground mines compared with surface mines.

The amount of coal recovery ranges over two orders of magnitude, from 11,794 to over 2,177,280 t (13,000 to over 2,400,000 st), with a median of 235,872 t (206,000 st). One measure of success of a remining operation is the amount of coal produced. This is because without the revised discharge standards, most mine operators are not willing to risk becoming responsible for perpetual treatment, and therefore the coal resource would go unused. This is especially true if the site is reclaimed through the abandoned mine lands program.

Key data of a remining permit are the water quality and flow of the preexisting mine drainage discharges. Alternative effluent standards are set based on those contaminant loading rates. Because water quality and flow data distributions are unpredictable and commonly asymmetrical, they are analyzed using exploratory data analysis and order statistics by the PADER. The performance of the remining operation in terms of additional mine water contamination is based on the results of these statistical analyses. Acidity loads from these sites range over four orders of magnitude, with a median of 20.3 kg/d (44.8 lb/d). Iron loads range over five orders of magnitude with a median of 0.54 kg/d (1.2 lb/d). Sulfate loads range over four orders of magnitude. The higher contaminant loading rates are directly related to the higher flowing mine discharges.

More than one abatement technique is usually employed during remining to abate or diminish the contaminant load. The most common techniques employed are regrading of dead spoils, underground mine daylighting, and spoil revegetation. Roughly one-third of the operations have alkaline addition and hydrologic controls as part of the abatement plan. The most common alkaline addition materials are limestone or dolostone, primarily because of their widespread availability and low cost. The pit floor is the most common location for the alkaline addition placement.

The estimated abatement costs range widely, from $0 to $4 million total cost. The average estimated cost per unit area is roughly twice the normal bond rate. Actual cost per metric ton of coal produced is also an indicator of the efficiency of the operation. However, estimated abatement costs exceeding $1.36 per metric ton ($1.50 per short ton) of coal may be somewhat inflated. At that cost level, based on the author's experience, it probably is not truly economically feasible to mine.

Experience indicates that the cost to treat the discharges to meet conventional effluent standards is not a fiscally viable option for the vast majority of the sites. The median cost of $5.90 per metric ton ($6.50 per short ton) to treat the discharges for 50 years illustrates this point. Projected costs at approximately 90% of the sites exceed $1.36 per metric ton ($1.50 per short ton) of coal.

Virtually every aspect of the Pennsylvania remining program indicates that it has been and continues to be successful, with only a few minor drawbacks. To date, the remining program in Pennsylvania has been successful in the permitting for reclamation of approximately 1,619 ha (4,000 acres) and has led to the production of over 32 million t (36 million st) of coal from areas deemed by many as "untouchable" under pre-remining regulations. The abatement techniques employed are geared toward reduction or elimination of mine discharge contaminant load. Estimated cost of abatement implementation is highly subjective and, experience indicates, often artificially high.

Data Analyses

To determine characteristics of hydrologic data from coal mines, the data were analyzed using several statistical techniques. Testing for normal distribution using the skewness and chi-square tests indicated that water quality and flow rate data tend to be nonnormally distributed. Remining appears to increase the tendency of these data to be nonnormally distributed, especially during the first few years after reclamation. The hydrologic data are commonly skewed to the right (the lower values). These trends are similar to those observed by other researchers for natural and degraded ground and surface waters. Trends exhibited by the skewness and chi-square tests indicate that flow is the dominant factor for determining the contaminant load rate.

Graphical analyses (notched box-and-whisker plots) indicate that underground mine discharges tend to be more severely degraded in terms of contaminant concentration than surface mine discharges in the remining data set. This is caused, in part, by differences in the ground water flow regime of the two mine types. Increased exposure of alkaline materials to ground water caused by surface mining also may be a factor in the differences in water quality. Flow rate has a strong influence on the contaminant load, although concentration can also be a significant influence.

Spearman's rank correlation analyses conducted on the hydrologic data illustrate that flow rate is more commonly strongly correlated with contaminant load than concentration. Therefore, if the discharge flow rate can be reduced by mining or reclamation practices, the mine operator may be able to virtually guarantee a reduction of the contaminant load. A flow reduction may be achieved by diversion or exclusion of ground water from adjacent areas from the spoil and/or by reducing surface infiltration.

The testing results indicate that the optimum baseline sampling duration of the three intervals analyzed (6, 9, and 12 months) is at least 1 complete year. With an entire year's worth of data, both wet and dry periods will be included in the discharge characterization, which will minimize the possibility of bias. Hornberger and others (1990) similarly concluded that an entire year is needed to characterize AMD discharges.

Time-consistent sampling (on a monthly, semi-monthly, or weekly basis) should prevent either wet or dry periods from being overemphasized.

Determination of the optimum baseline sampling frequency illustrates that the highest sampling rate for each SES was consistently not the best rate to characterize the discharges. A comparison of the results using the two PADER triggering methods exhibited some dissimilarities. However, a time-consistent sampling rate, when sampling is on a monthly basis, appears to adequately characterize the contaminant load at minimum cost. If cost of monitoring is not a consideration, a semi-monthly rate should be at least equally as good as a monthly rate, if the sampling is consistent with regard to time (e.g., samples collected on the first and third Monday of the month).

Analysis of the contaminant concentrations, loading rates, and flow rates of mine discharges using several methods indicates that Pennsylvania's remining program is successful from the standpoint of preventing additional ground and surface water degradation. The overwhelming majority of the discharges have post-remining contaminant loads of acidity and iron that are equivalent to or significantly less than the pre-remining levels. Short-term changes (less than 1 year) in flow and/or concentration are the primary reasons that significant degradation appears to have occurred at a number of discharges.

Reduction of Discharge Flow

When any of the methods of analyses indicate that a significant change in contaminant load occurred, changes in the discharge flow rate is by far the most common reason. Concentration is a possible factor in some cases. Concentration may play a somewhat stronger role when a significant *increase* in contaminant load is indicated than when a significant decrease is indicated.

Because of the strong control that the mine discharge flow rate exerts on the corresponding contaminant load, if flow can be reduced through mining and/or reclamation practices, the probability that the remining operation will not incur treatment liability on a long-term basis is greatly increased. With this knowledge, mine operators may be more willing to enter into remining permits and regulating agencies may be more receptive to issue them.

Practices to reduce discharge flow can be incorporated into the permit application abatement plan. Flow reduction can be achieved by exclusion or diversion of ground and surface water away from the reclaimed site. Methods that decrease surface water recharge include installation of diversion ditches, capping the site with a low-permeability material, spoil regrading, and revegetating. Abandoned sites, prior to remining, commonly have unreclaimed pits and closed-contour depressions in the poorly sorted spoil that serve as recharge zones for significant quantities of infiltrating surface water. For many abandoned surface mines, the act of regrading and revegetating spoil will significantly reduce surface water infiltration and increase runoff just by the elimination of these recharge zones. This may be the most viable option; it is the least expensive method of reducing surface recharge because it must be performed to satisfy the reclamation requirements.

Methods for decreasing ground water recharge to the spoil include installation of drains and/or grout curtains near the final highwall, drains running the length of the pit floor, and horizontal free-draining dewatering wells, and sealing of adjacent underground mine entry ways exposed during mining. Where the remining is daylighting of underground mines, sealing of entry ways may be the least expensive and most viable option. When abandoned surface mines are remined, installing the highwall drain may be the most viable option, if sufficient grade can be achieved to allow a free-draining, low-maintenance system.

Future Remining

The three statistical methods for determining changes in discharge contaminant load yielded similar results. Each of these methods would be applicable for use in a remining program, if placed in a framework similar to the one PADER currently uses.

Although this study was conducted exclusively in Pennsylvania, similar remining programs in other Appalachian States should be at least as successful in terms of contaminant load reduction. The geologic and hydrologic conditions in these other States are similar enough to those of the western Pennsylvania coal fields to facilitate similar results.

ACKNOWLEDGMENTS

REFERENCES

Aljoe, W. W., and J. W. Hawkins. Application of Aquifer Testing in Surface and Underground Mines. Paper in Proceedings NGWA—FOCUS Conference on Eastern Region Ground Water Issues. NGWA, 1992, pp. 541-555.

Berryman, D., B. Bobee, D. Cluis, and J. Haemmerli. Nonparametric Tests for Trend Detection in Water Quality Time Series. Water Resour. Bull., v. 24, No. 3, 1988, pp. 545-556.

Commonwealth of Pennsylvania. Pennsylvania Bulletin. V. 15, No. 26, June 29, 1985, pp. 2377-2389.

Davis, J. C. Statistics and Data Analysis in Geology. Wiley, 1986, 646 pp.

Gibbons, R. D. A General Statistical Procedure for Ground-Water Detection Monitoring at Waste Disposal Facilities. Ground Water, v. 28, No. 2, 1990, pp. 235-243.

Harris, J., J. C. Loftis, and R. H. Montgomery. Statistical Methods of Characterizing Ground Water Quality. Ground Water, v. 25, No. 2, 1987, pp. 185-193.

Hawkins, J. W., and W. W. Aljoe. Hydrologic Characteristics of a Surface Mine Spoil Aquifer. Paper in Proceedings of the Second International Conference on the Abatement of Acidic Drainage (Montreal, Quebec, Canada). MEND, 1991, pp. 47-68.

Helsel, D. R. Advantages of Nonparametric Procedures for Analysis of Water Quality Data. Hydrol. Sci. J., v. 32, No. 2, 1987, pp. 179-190.

_____. Mine Drainage and Rock Type Influences on Eastern Ohio Stream Water Quality. Water Resour. Bull., v. 19, No. 6, 1983, pp. 881-887.

Hirsch, R. M., and J. R. Slack. A Nonparametric Trend Test for Seasonal Data With Serial Dependence. Water Resour. Res., v. 20, No. 6, 1984, pp. 727-732.

Hornberger, R. J., M. W. Smith, A. E. Friedrich, and H. L. Lovell. Acid Mine Drainage From Active and Abandoned Coal Mines in Pennsylvania. Ch. in Water Resources in Pennsylvania: Availability, Quality and Management, ed. by S. Majumdar, E. Miller, and R. Parizek. The PA Acad. Sci., 1990, pp. 433-451.

Kohlmann Ruggiero Engineers, P.C. Preliminary Engineering Cost Manual for Development of BPJ Analyses for Coal Remining, Four Case Histories in Pennsylvania. Prepared for U.S. EPA and PA Dep. Environ. Resour., 1986.

_____. Development of a Waste Water Treatment Cost Analysis Module for the Coal Remining BPJ Computer Software Package. Prepared for U.S. EPA, 1990, 65 pp.

Lindorff, D. E. Hydrogeology of Surface Coal Mines in Illinois. Paper in Proceedings of the Symposium on Surface Mining, Sedimentology and Reclamation (Lexington, KY). A.S.S.M.R., 1980, pp. 33-40.

Lineberry, G. T., K. F. Unrug, and D. E. Hinkle, Estimating Secondary Mining Potential of Inactive and Abandoned Appalachian Highwalls. Int. J. Surf. Min. Reclam., v. 4, 1990, pp. 11-19.

McGill, R., J. W. Tukey, and W. A. Larsen. Variations of Box Plots. Am. Stat., v. 32, No. 1, 1978, pp. 12-16.

Montgomery, R. H., J. C. Loftis, and J. Harris. Statistical Characteristics of Ground-Water Quality Variables. Ground Water, v. 25, No. 2, 1987, pp. 176-184.

Norcliffe, G. B. Inferential Statistics for Geographers. Wiley, 1977, 272 pp.

O'Steen, W. N., and H. W. Rauch. Effects of Coal Mining on Ground-Water Quality in West Virginia. Paper in Proceedings of the Symposium on Surface Mining, Hydrology, Sedimentology and Reclamation (Lexington, KY). A.S.S.M.R., 1983, pp. 355-363.

Pennsylvania Department of Environmental Resources, Pennsylvania State University, and Kohlmann Ruggiero Engineers. REMINE. 1988.

Phelps, L. B., and M. A. Thomas. Assessment of Technological and Economical Feasibility of Surface Mine Reclamation, Stage One Research. Draft final Rep. prepared for PA Dep. Environ. Resour., 1986.

Razem, A. C. Ground-Water Hydrology Before, During, and After Coal Strip Mining of a Small Watershed in Coshocton County, Ohio. U.S. Geol. Surv., Water-Resour. Invest. Rep. 83-4155, 1983, 39 pp.

Reed, L. A. Effects of Strip Mining the Abandoned Anna S Mine on the Hydrology of Babb Creek, Tioga County, Pennsylvania. U.S. Geol. Surv., Water Resour. Invest. Rep. 80-53, 1980, 41 pp.

Richardson, A. R., and M. T. Dougherty. Feasibility Study, Deer Park Daylighting Project. U.S. EPA, EPA-600/2-76-110, 1976, 76 pp.

Smith, M. W. Establishing Baseline Pollution Load From Preexisting Pollutional Discharges for Remining in Pennsylvania. Paper in Mine Drainage and Surface Mine Reclamation. II. Mine Reclamation, Abandoned Mine Lands and Policy Issues. USBM IC 9184, 1988, pp. 311-318.

Snedecor, G. W., and W. G. Cochran. Statistical Methods. IA State Univ. Press, Ames, IA, 1971, 593 pp.

Tukey, J. W. Exploratory Data Analysis. Addison-Wesley Publ., Reading, MA, 1976, 638 pp.

Velleman, P. F., and Hoaglin, D. C. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Wadsworth, Belmont, CA, 1981, 354 pp.

APPENDIX A.—SUMMARY OF LOADING DATA[1]

| Site | Pre-remining | | | | Post-remining | | | |
| | n | Load, kg/d | | | n | Load, kg/d | | |
		Acidity	Iron	Sulfate		Acidity	Iron	Sulfate
1	17	0.2	0.00	3.8	13	0.09	0.00	1.2
2	21	565.0	42.93	1,573.3	41	454.80	40.52	1551.8
3	22	20.5	2.25	41.0	14	1.23	0.09	3.5
4	10	10.7	0.54	20.8	45	19.92	3.14	103.4
5	38	3.7	0.03	20.5	19	4.31	0.02	36.1
6	21	27.1	0.38	84.3	16	41.16	0.63	185.7
7	31	612.4	7.87	666.6	11	678.32	8.71	901.6
8	10	163.5	20.48	430.0	40	33.84	5.68	340.1
9	4	304.4	14.72	618.8	16	77.67	5.56	268.1
10	9	0.9	0.18	23.9	33	0.70	0.15	12.8
11	6	145.0	5.49	512.8	17	14.58	0.38	48.3
12	12	38.3	0.21	NAp	63	4.70	0.09	56.6
13	11	0.4	0.01	NAp	56	0.50	0.01	27.1
14	28	0.1	0.01	1.0	46	0.25	0.22	6.7
15	9	4.4	0.10	5.7	24	2.47	0.02	11.3
16	3	128.4	74.05	360.7	71	123.58	54.73	320.6
17	26	8.2	0.13	57.3	43	0.13	0.10	45.2
18	24	0.8	0.23	73.6	33	2.53	0.10	112.3
19	18	0.7	0.01	3.3	36	5.11	0.06	23.9
20	8	52.5	0.88	178.4	7	0.02	0.00	2.4
21	16	9.5	0.18	22.6	21	6.22	0.20	32.9
22	28	79.1	1.35	345.5	10	88.67	1.81	457.2
23	8	51.1	2.05	136.6	25	30.99	2.25	117.1
24	18	9.7	1.18	11.7	12	0.00	0.00	0.0

NAp   Not applicable.
[1] All data are median values, except for the number of samples (n).

APPENDIX B.—PLOTS OF ACIDITY, IRON, AND SULFATE LOAD MEDIANS
OF UNDERGROUND AND SURFACE MINES BEFORE AND AFTER REMINING

Figure B-1
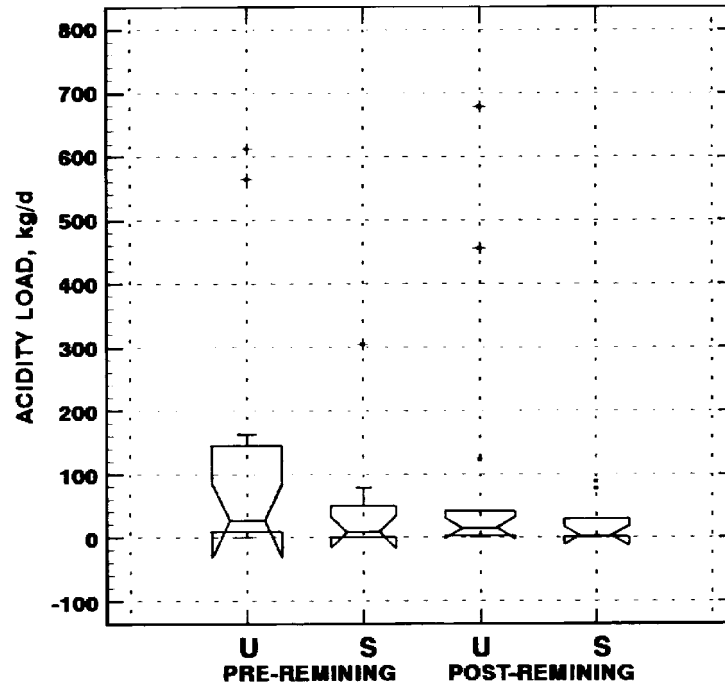
Acidity load medians of undergroundU) and surface (S)
mines before and after remining .

Figure B-2
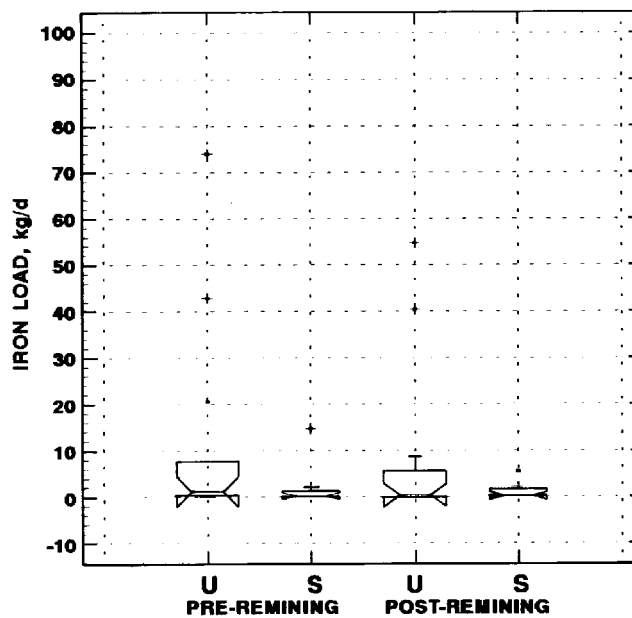
Iron load medians of underground (U) and surface (S)
mines before and after remining.

Figure B-3

Sulfate load medians of underground (U) and surface (S)
mines before and after remining.

APPENDIX C.—SUMMARY OF ANALYSES OF IRON AND SULFATE LOADS BEFORE AND AFTER REMINING

Table C-1.—Analyses of iron load before and after remining

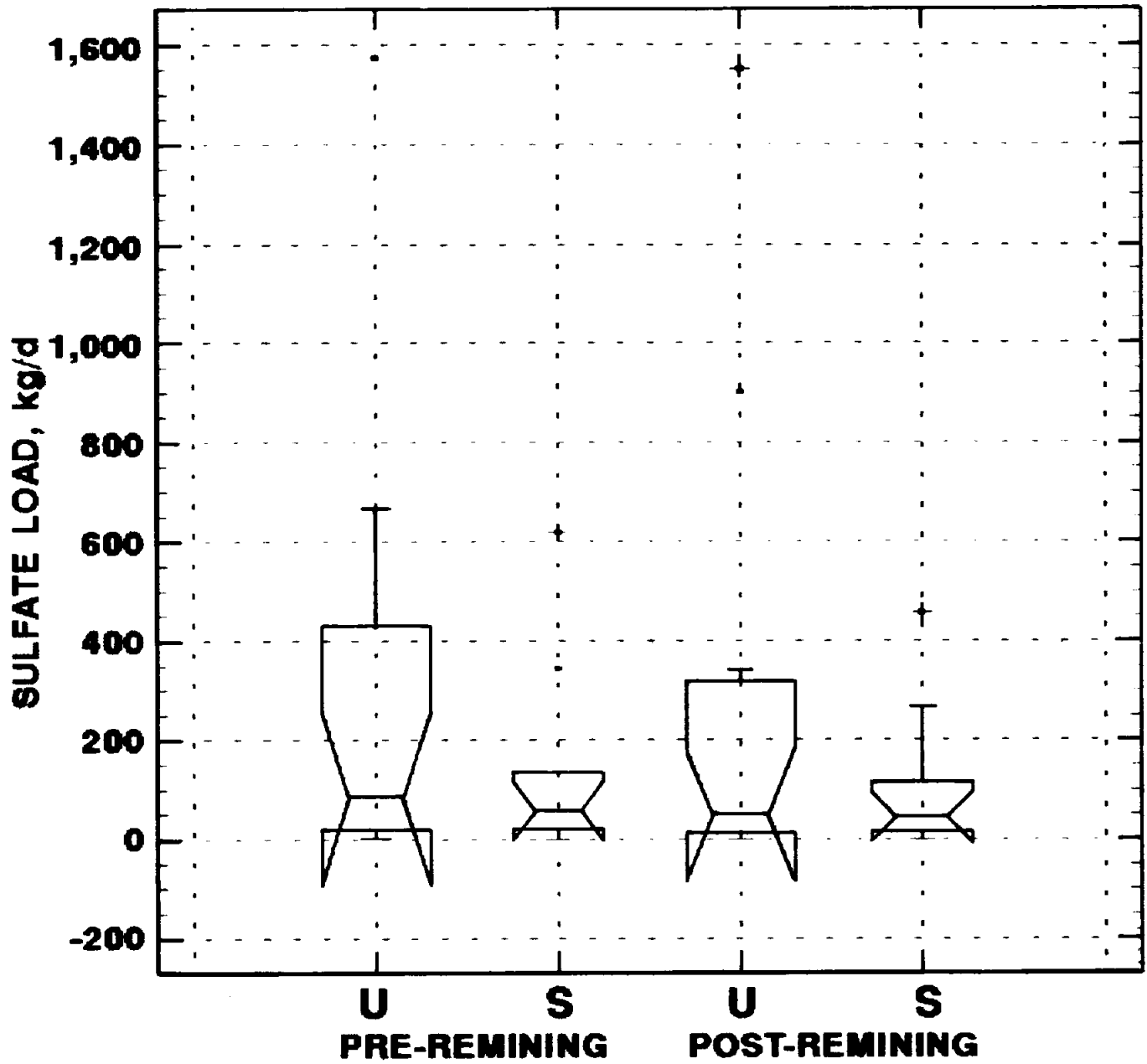| Site	PADER system type		Mann-Whitney U test	Gibbons NUPL method			
				2 samples		3 samples	
	1	2		5%	1%	5%	1%
1	Y	N	N	N	N	N	N
2	N	N	N	N	N	N	N
3	N	N	N	N	N	N	N
4	Y	Y	Y	Y	X	Y	Y
5	N	N	N	N	N	N	N
6	N	N	Y	N	N	N	N
7	N	N	N	N	N	N	N
8	N	N	N	Y	X	Y	Y
9	N	N	Y	X	X	N	X
10	N	N	N	N	N	N	N
11	N	N	Y	N	X	N	X
12	N	N	N	N	N	N	N
13	N	N	N	N	N	N	N
14	N	Y	Y	Y	Y	Y	Y
15	N	N	N	N	X	N	N
16	N	N	Y	X	X	X	X
17	N	N	Y	N	N	N	N
18	N	N	N	N	N	N	N
19	Y	Y	Y	Y	Y	Y	Y
20	N	N	N	N	X	N	N
21	N	N	N	N	N	N	N
22	N	N	Y	N	N	N	N
23	N	N	Y	N	X	N	N
24	N	N	N	N	N	N	N

NUPL   Nonparametric upper prediction limits.
PADER   Pennsylvania Department of Environmental Resources.

Note.—N (no) indicates treatment was not initiated under the PADER system type 1, the post-remining median did not exceed the pre-remining 95% confidence interval about the median under the PADER system type 2, the Mann-Whitney U test indicated no significant diffrence at the 5% confidence level, and the Gibbons NUPL method did not have two or three consecutive samples exceeding the pre-remining maximum. Y (yes) indicates that treatment was initiated under the given system. X indicates data were not available or were insufficient to complete the analyses.

Table C-2.—Analyses of sulfate load before and after remining

Site	PADER system type		Mann-Whitney U test	Gibbons NUPL method			
				2 samples		3 samples	
	1	2		5%	1%	5%	1%
1	Y	N	Y	N	N	N	N
2	N	N	N	N	N	N	N
3	N	N	N	N	N	N	N
4	Y	Y	Y	Y	X	Y	Y
5	N	N	N	Y	Y	N	N
6	N	N	Y	Y	Y	Y	Y
7	N	N	N	N	N	N	N
8	Y	N	Y	Y	X	Y	Y
9	N	N	Y	X	X	N	X
10	N	N	N	N	N	N	N
11	N	N	Y	N	X	N	X
12	X	X	X	X	X	X	X
13	X	X	X	X	X	X	X
14	N	Y	Y	Y	Y	Y	Y
15	N	N	N	Y	X	N	N
16	N	N	N	X	X	X	X
17	N	N	Y	N	N	N	N
18	N	N	Y	N	N	N	N
19	Y	Y	Y	Y	Y	Y	Y
20	N	N	N	N	X	N	N
21	Y	N	Y	Y	Y	Y	Y
22	N	N	N	N	N	N	N
23	N	N	Y	Y	X	N	N
24	N	N	N	N	N	N	N

NUPL   Nonparametric upper prediction limits.
PADER   Pennsylvania Department of Environmental Resources.

Note.—N (no) indicates treatment was not initiated under the PADER system type 1, the post-remining median did not exceed the pre-remining 95% confidence interval about the median under the PADER system type 2, the Mann-Whitney U test indicated no significant difference at the 5% confidence level, and the Gibbons NUPL method did not have two or three consecutive samples exceeding the pre-remining maximum. Y (yes) indicates that treatment was initiated under the given system. X indicates data were not available or were insufficient to complete the analyses.