A statistical evaluation of remining abandoned coal mines to reduce the effluent contaminant load

Jay W. Hawkins
US Bureau of Mines, Pittsburgh, Pa., USA

ABSTRACT: Ascertaining changes in contaminant load of mine discharges is crucial for determining the success of coal remining operations in terms of pollution load reduction. The method of data analysis used to make these determinations is equally important. Pre- and post-remining water quality and flow rates of 57 discharges from 24 remining operations in western Pennsylvania were analyzed using exploratory data analysis, the Mann-Whitney U test, and nonparametric upper prediction limits. The three methods of data analysis yielded similar results. The analyses indicate that a majority of the sites exhibited either no change or a significant decrease in pollution rate because of remining. The discharge flow rate was determined to be the dominant controlling factor when the post-remining pollution load was observed to be significantly different (better or worse) than the pre-remining load. Concentration was a subordinate factor. The relative importance of concentration for load determination tended to increase when the post-remining load rate significantly exceeded the pre-remining rate. When the analyses indicated that mine discharges have been degraded from remining, this was generally caused by short-term changes in flow and/or concentration shortly after reclamation.

1 INTRODUCTION

Remining, in this context, is the surface mining of abandoned surface and/or underground mines that originally created and continue to possess discharges that fail to meet the applicable effluent standards. Under a remining program, an operator can affect these sites without assuming legal responsibility for treatment of the previously degraded water, as long as the discharging waters are not further degraded by the additional mining. If the water is additionally degraded, the required treatment is based on site-specific pre-remining pollution levels and not legislatively-promulgated effluent standards. In Pennsylvania, to establish pre-remining pollution load levels and for later determination if additional degradation has occurred, the mine operator must collect a consecutive series of pre-remining discharge water samples along with flow rate measurements. Remining pollution (baseline) loading rate effluent standards (e.g., kilograms or pounds of pollutant per day) are established based mainly on the analyses of these data. However, the baseline loading rate is also somewhat dependent on the strength of the pollution abatement plan as well as the economics of treating the water to conventional effluent standards. In order to receive a remining permit, the operator must demonstrate that there is a potential to improve the water quality. If, after remining, the pollution loading rates are no worse than the pre-remining loading rates and all other physical and temporal reclamation requirements are satisfied, discharge monitoring ceases and the operator's performance bonds are released.

In the absence of a remining program, most mine operators avoid abandoned sites with minable coal reserves and pre-existing pollutional discharges. Under these circumstances, the mine operator would be liable for treatment in perpetuity for all discharges that failed to meet the applicable statutory effluent standards. This is the case, even if the pre-existing discharges were caused by previous mining unrelated to the subsequent operation and the water quality was improved by the remining.

For this project, 105 remining permits from the bituminous district of Pennsylvania were reviewed. Twenty-four completed operations were selected from those permits for the remining impact study. Figure 1 is a map of Pennsylvania showing the location of these 24 sites. These sites were selected because they had been completed (backfilled to rough grade) and possessed a minimum of one year of post-backfilling water quality and flow data.

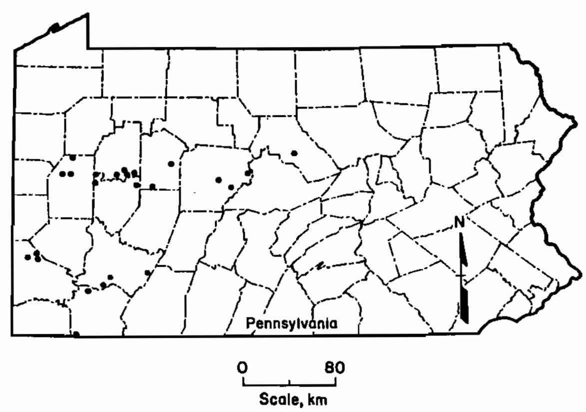

Figure 1. Location map of the study sites.

To determine the effectiveness of remining in terms of reduction in pollution load, 57 discrete discharges from the 24 remining operations were analyzed. Table 1 is a summary of the data from these sites. The pre- and post-remining data were analyzed using a system employing exploratory data analysis (schematic summary), which is the method presently used by the Pennsylvania Department of Environmental Resources (PADER), the Mann-Whitney U test, and a method of nonparametric upper prediction limits. The test results were also used to assess the applicability of each of the analytical methods to these types of data.

2 BACKGROUND OF THE PENNSYLVANIA REMINING PROGRAM

The PADER has been issuing remining permits in the bituminous coal fields since 1983. Remining regulations were officially appended to the Pennsylvania Surface Mining Conservation and Reclamation Act by Act 158 of 1984. These changes were subject to approval by the U.S. Environmental Protection Agency and the Office of Surface Mining. In its present form, the remining program in Pennsylvania (Subchapter F program) was made part of the permitting process in 1986 by use of a permit application module (module 26) and a series of standard special conditions that are appended to the permit.

The PADER remining effluent standards are based mainly on pre-remining water quality and flow rates. Remining effluent standards are set as baseline pollution loading rates in units of kilograms or pounds of pollutant per day. This is in contrast to the usual pollutant effluent limits, which are in units of pollutant concentration (e.g. milligrams per liter), as set by 40 CFR part 434.30 on the Federal level or 25 PA. Code Chapter 87, Section 87.102 by the Commonwealth of Pennsylvania.

Pennsylvania was selected for this study because it is the only state in the Appalachian region that has a fully operational, well-documented, and time-tested remining program. No other state remining program approaches the magnitude of this program in terms of the number of permits issued or in the background and operation documentation. Pennsylvania has been issuing remining permits for nearly 10 years; as of March 1992, over 110 remining permits had been issued. Some of the other Appalachian states have issued a couple of remining permits, while others have yet to begin issuance (Veil, 1993).

3 RELATED STUDIES

Crucial to any remining program are the methods of analysis that are 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 groundwater 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 groundwater 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.

Lineberry and others (1990) developed a mathematical procedure for evaluating the potential for secondary mining (remining) of abandoned exposed highwalls in Appalachia. They estimated that approximately 70% of the 21.8 billion tonnes (24 billion tons) within 60 meters (197 feet) of exposed highwalls and 90.7 billion tonnes (100 billion tons) within 180 meters (591 feet) of exposed highwalls are recoverable by present remining techniques, mainly highwall and auger mining.

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, Maryland. They collected water quality data before and after daylighting in order to evaluate its impact. A preliminary final report concerning this project was completed by Ackerman and others (undated). They observed that the post-remining pollutant loads exhibited no significant difference from the pre-remining levels, although the post-remining pollutant concentrations exhibited seasonal fluctuations that were significantly higher than 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 to pre-remining conditions.

Reed (1980) analyzed the effect of daylighting on a 344 hectare (850 acre) underground mine in Tioga County, Pennsylvania. 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.

Smith (1988) determined temporal and seasonal characteristics on the flow rate, acidity concentration and load of mine discharges from three hydro-geologically different mines in Pennsylvania. These discharges were analyzed to assess optimal sampling frequency and duration to accurately characterize the pre-remining baseline pollution load. Acidity baseline loading rates appeared to be mainly controlled by flow rate. He recommended a minimum of one year of pre-remining flow and water quality sampling with a frequency that was at least monthly. He also suggested that, depending on the type of discharge, this frequency may need to be significantly increased in order to characterize short-term high or low loading periods. Regardless of the frequency, sampling must be consistent in order not to overemphasize any one period.

Most studies pertaining to the effect of coal mining on water quality have been limited to its impact on previously-unmined areas. 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 exhibited neither a normal nor 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: data transformations are not required, the tests can be performed even if normality of data sets is not achievable, greater power in highly-skewed data sets, central tendency comparisons are made on the median rather than the mean, below detection-limit data can be easily incorporated without bias. The potential bias is further reduced by the use of the data median which is less sensitive than the mean to a few extreme 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 greater than 100 mg/L in shallow ground water after 20 years (O'Steen and Rauch, 1983). 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."

4 DATA ANALYSES AND RESULTS

In general, hydrologic data are 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) and Hawkins (1993) 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 (right). To determine the hydrologic impact of remining, the pre-remining pollutant concentration, discharge rate, and pollutant load will be compared to post-remining data utilizing an exploratory data analysis system employed by the PADER, the Mann-Whitney U test (a nonparametric form of the Student's t test), and a nonparametric method of upper prediction limits developed by Gibbons (1990). Secondarily, this paper will ascertain the comparative applicability of these methods for detecting changes in the mine discharge quality.

4.1 PADER system (exploratory data analyses)

In Pennsylvania, pre-remining pollution loading rates are set during the remining permit application process based on a consecutive series of pre-remining water quality samples and measured discharge flow rates. Initially, a minimum of 6 consecutive 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 may have less than 6 samples (Table 1). The PADER now requires 12 consecutive months of data or at least data collected from February through October. Baseline loading rate calculation must use consecutive monthly sampling from an entire water year or water years. Partial water year data generally cannot be used. In theory, 12 consecutive monthly samples will include both low flow and high flow periods in the background data, which will more accurately characterize the preexisting discharges. At a minimum, the water samples must be analyzed for concentrations of alkalinity, acidity, total iron, total manganese, total aluminum, sulfate, total suspended solids and pH. The pollutants are generally reported in units of milligrams per liter (mg/L), except for pH, which is in standard units. Discharge flow rate, usually reported in gallons per minute (gpm), is gaged by a variety of means including the use of a weir, flume, cross sectional area with a flow meter, or bucket-stopwatch method. The flow rate was subsequently converted to liters per minute (L/m). The weir appears to be the most common method used for mine discharges over 38-76 L/m (10-20 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 (lbs/day) are calculated from the flow and concentration data. For this study, the data were converted to kilograms per day (kg/day).

The pre-remining water quality data are analyzed using some basic exploratory data analyses (schematic summary) and nonparametric statistics (PADER, 1988). The results for each discharge or hydrologic unit are presented in tables containing the loading data range, the median, the first and third quartiles, the approximate 95% tolerance limits (depth of 32nds values or C spread), and 95% confidence interval about the median for each of the regulated pollutants. Table 2 is an example of these results. Tukey (1976) and Velleman and

Table 1. Summary of remining data. All data are median values, except n which is the number of samples.

	PRE-REMINING					POST-REMINING				
Site	n	flow L/m	acidity mg/L	Fe mg/L	SO_4 mg/L	n	flow L/m	acidity mg/L	Fe mg/L	SO_4 mg/L
1	17	8	19	0.2	339	13	4	30	0.2	420
2	21	1181	321	24.1	814	41	1283	261	20.5	852
3	22	23	511	62.0	1132	14	4	512	36.0	1089
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	1020	11.3	1077	11	466	850	12.8	1087
8	10	204	777	99.3	1695	40	117	294	37.6	2189
9	4	261	1447	58.1	2671	16	95	742	54.7	2230
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	1430	71	148	541	218.5	1202
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	1816

Hoaglin (1981) present detailed explanations of the method by which these values are calculated and what they represent. Under the PADER system, there are four mechanisms by which treatment of a discharge can be triggered (initiated) using the pre-remining water quality statistical data in Table 2.

The main triggering method under this system requires a series of 6 consecu- tive samples to exceed the upper bound of the approximate 95% tolerance limits (item 4 in table 2). Discharge sampling, during and after mining, is performed on a monthly basis until the reclamation performance bonds have been released (generally Stage II bonds). If, during this period, two consecutive samples exceed the approximate 95% tolerance limits for any of the specified pollutants, this immediately triggers weekly sampling. If 4 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 at any time during the weekly sampling, then monthly monitoring resumes and treatment is not required at that time. If, while treating, two consecutive weekly samples drop below the 95% tolerance limits, then treatment is likewise suspended.

A second mechanism for treatment initiation is when statistical analysis indicates the during- or post-remining median pollutant load has been increased in comparison to the pre-remining median at the 5% significance level. This is determined by comparison of the 95% confidence interval about the median (Table 2, item 5) of the pre- and post-remining data. In this case, the median is calculated on a complete water year (October 1 through September 30) basis.

The third triggering method uses the same method of analyses. However, the median is determined for water-year periods No. 1 (October 1 through April 30) or No. 2 (May 1 through September 30). If either of these periods indicate the post-remining median is above the pre-remining median at the 5% significance level, this will likewise trigger treatment. This method is seldom used because separate parameters have to be calculated for each water year period prior to permit issuance. 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 pollution load is unrelated to the mining operation and is caused by factors beyond their control (e.g., adjacent unrelated mining operations or an extreme storm event) treatment of the discharge is not required.

Treatment standards are likewise based on the data contained in table 2. The treated average monthly discharge pollution load cannot exceed the pre-remining median and is calculated based on the samples collected weekly. The instantaneous maximum pollution load permitted is based on a "grab" sample and must be no greater than the upper quartile, ("HIGH" value of item 3 in table 2). Both parameters must be reported at least monthly.

If the pre-existing degraded discharges are physically encountered during remining, they must be treated to the conventional best-available-treatment effluent standards. Once these discharges are no longer physically encountered by the mining operation, the modified effluent standards, under the remining program, are reinstated. This generally occurs during the reclamation stages of a remining operation.

Table 3 summarizes the number of sites that triggered the PADER system for weekly sampling and subsequently treatment for acidity, iron, and sulfate. 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 nonparametric upper prediction limits, the discharges were analyzed separately in an attempt to provide a more detailed assessment of the effects of remining.

Two of the sites did not have data for pre-remining sulfate concentration, therefore sulfate "standards" were established for only 22 of the sites. Sulfate, although not a regulated contaminant for surface water, was included in these analyses because it is generally a conservative indicator of acid mine drainage production. Increases in the sulfate content of mine water in the Appalachian coal mining region indicate an acceleration of metal sulfide, usually iron disulfide, oxidation. Sulfate ion (SO_4^{-2}) is released by this reaction. The sulfate is little affected by geochemical changes to the mine water and remains in solution to relatively high concentrations, governed primarily by the solubility of gypsum ($CaSO_4*2H_2O$).

Weekly water sampling was initiated at least once during the post-remining period by acidity for 10 of 24 sites. Three of these sites subsequently triggered treatment. Iron was the triggering contaminant for weekly sampling in 5 out of 24 sites. Of these 5 sites, 3 secondarily triggered discharge treatment. Of the 3 incidences of discharge treatment for acidity and iron, two occurred concurrently, therefore treatment was actually initiated at least once on 4 sites. The

Table 2. Example of baseline pollution load summary used by the PADER for remining permitting.

Mine ID: Mine Name: Hydrologic Unit ID:

		LOADING IN POUNDS PER DAY[1]			
PARAMETER:		FLOW (gpm)	ACIDITY	IRON	SULFATE
NUMBER OF SAMPLES (N):		43	43	43	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

1 - These units are as used for remining permitting by the PADER.

Table 3. Summary of weekly sampling frequency and discharge treatment triggering after remining. "Yes" indicates that triggering occurred at least once to one or more discharges/hydrologic units for that site. "No" indicates triggering never occurred.

	Acidity		Iron		Sulfate	
	Yes	No	Yes	No	Yes	No
Triggered Weekly Sampling	10	14	5	19	11	11
Triggered Treatment	3	21	3	21	5	17

elevated pollution loads that triggered treatment were in all cases transient events. The higher loads generally started shortly after reclamation (less than one year) and in all cases declined to within standards in a short period (less than 6 months). Sites 4, 8, and 19 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 the site was backfilled.

Sulfate would have triggered weekly sampling, if it was an 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 initiate 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 mine drainage 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 will reduce the acidity concentration and raise 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 given sulfate and calcium levels. With this in mind, it is possible for mine water to exhibit significant increases in sulfate without corresponding increases in acidity or iron.

The results shown in table 3, created using the first triggering method of the PADER system, illustrate that discharge treatment is seldom incurred. A review of the cases where treatment was initiated indicate that it was of a ephemeral nature and occurred most often shortly after reclamation.

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 to the median for the pre-remining data. Lack of overlap of 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 or water year period is to insure that the data used in the comparison are not biased by most of the sampling occurring during a particularly low or high flow period. The large number of possible water year and water year periods for 24 sites with 57 discharges precluded the strict adherence to this part of the PADER system. Instead, the median of the data for the entire post-remining period was compared with the pre-remining data. Any sampling bias of the data is minimized because each of the post-remining data sets include 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 discharging mine waters (Table 4). The median acidity and iron loads for 21 out of 24 sites was equal to or below the pre-remining levels at the 5% significance level. The one site where acidity was above pre-remining levels coincided with one of the iron excursions. Therefore, there were a total of 3 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 for acidity from 7 of 24 sites and for iron from 4 of 24 sites were below the pre-remining median. This indicates that nearly a third of the sites had a statistically significant improvement in acidity load. For three of the sites, the acidity and iron excursions coincided, making a total of 8 sites with a decrease in contaminant load. Of the 11 incidences where acidity and iron loads were significantly below the pre-remining median, 7 likewise exhibited a significant flow reduction. However, concentration appeared to have contributed in 6 of these excursions. The median concentration dropped by a factor of two or more for these 6 excursions.

Nearly 13% (3 of 24) of the sites exhibited a median acidity and iron loads above the pre-remining median, indicating an apparent significant increase in contaminant load. Of the 4 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 must also play 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 exhibits similar results as acidity and iron loads. Flow rate changes were a significant factor for 3 out of 4 of the 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 2 out of the 3 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 pre-remining 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. Analysis of the adequacy of the pre-remining sampling frequency and duration is beyond the scope of this paper, but will be the focus of future efforts.

Table 4. Summary of a comparison of the post-remining mine site median contaminant loads to the pre-remining the median at the 5% significance level.

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

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

4.2 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 where 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 median, rather than the mean, of the data sets were 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 mine site 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 may promote a more accurate assessment of hydrologic changes caused by remining.

The MW test results (table 5) 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. The number of discharges above the 5% significance level (indicating degradation) is lower for acidity and iron loads than the corresponding concentrations. This appears to be related to the strong influence that the discharge flow rate has on the contaminant load, as observed by Smith (1988) and Hawkins (1993). The acidity and iron loading MW test results are very similar to those exhibited by the flow rate and dissimilar to the corresponding concentration.

The number of discharges exhibiting increased sulfate concentration and load from remining is substantially higher than those 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 changes to ground water flow paths and contacted material caused by mining and reclamation, as previously discussed. This does not necessarily imply that remining has caused a pollution 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 compared to acidity or iron.

Table 5. Summary of the Mann-Whitney U Test comparing the post-remining to the pre-remining data. Below represents the number of post-remining parameter medians that were below the corresponding pre-remining median at the 5% significance level. Above means the median was above at the 5% level and within indicates that the data did not exceed either the upper or lower 5% significance level.

	Below	Above	Within
Acidity Concentration	24	7	26
Acidity Load	17	4	36
Iron Concentration	18	9	30
Iron Load	17	5	35
Sulfate Concentration	9	18	25
Sulfate Load	14	10	28
Flow Rate	16	4	37

4.3 Nonparametric upper prediction limits

A method of nonparametric upper prediction limits (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 out 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 3 and 5). This method was originally developed to determine degradation of 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 for the pre- and post-remining comparison of hydrologic data from individual discharges of remining sites.

The remining data were analyzed by use of tables and, as required, the equation derived by Gibbons (1990). The upper limit probability was established at the 5 and 1% significance levels (95 and 99% confidence, respectively). Comparisons were made on 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 mine site.

In order to achieve 5 and 1% significance levels with two post-remining resamplings and one 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 3 and 7, 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 8 at the 1% significance level with two resamplings. With three resamplings, the number of sites with insufficient background samples were reduced to 1 and 2 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 6). The number of discharges below the 5% significance level increased slightly, from 76 to 85%, when the number of resamplings were raised from two to three. However, at the 1% significance level, there was little change. The increase at the 5% level 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.

Of the three contaminant loadings, sulfate most often exceeded the 5 and 1% significance levels (table 6). This is the same general trend exhibited when sulfate data were analyzed using the PADER system and the MW tests. As previously discussed, greater range in sulfate concentrations could explain the higher rate of "failure" in terms of sulfate.

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) and Hawkins (1993). 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 to acidity and iron loads.

In total, 8 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 2 resamplings. With 3 resamplings, 5 and 4 sites exceeded the 5 and 1% significance level for acidity and/or iron, respectively.

4.4 Comparison of the methods of analysis

A site-by-site comparison of the three analytical methods (table 7) for acidity loads yielded generally similar results. Similar results were likewise exhibited in a comparison for iron and sulfate (not shown). An indication of degradation under one of the two Pennsylvania methods was generally reflected in the MW test and/or the NUPL method. However, there were a few sites where degradation was indicated by the MW test or the NUPL method that were not indicated by the Pennsylvania methods. This trend is partially caused by differences in the methods of data analysis. Under the Pennsylvania system, each site was analyzed as a single hydrologic unit. However, the MW test and the NUPL methods, each of the discharges were analyzed separately. An indication of degradation, as shown in table 7, using the MW test or the NUPL method indicates that one or more discharges from that site exceeds the applicable standard. On some sites (4, 14, and 16), one discharge indicated increased degradation from acidity, while the rest were within expected limits or were below. For this reason, acidity load for the corresponding site may show no change or a net decrease using the Pennsylvania system.

Mine sites with discharges exhibiting both decreases and increases in pollutant 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 (MW and NUPL) and the two Pennsylvania system methods, which look at the site as a single hydrologic unit.

The comparison of analytical methods suggest that the MW 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 the PADER system, either of these methods should adequately determine degradation, nondegradation, or improvement due to remining.

5 SUMMARY AND RECOMMENDATIONS

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

When any of the methods of analysis indicate that a significant change in terms of pollution load occurred, changes in the discharge flow rate is the most common reason. Concentration is a subordinate factor in some cases. Concentration may play a somewhat stronger role when a significant increase in pollution 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 pollution load, if flow can be reduced through mining and/or reclamation practices, the probability that the 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 conducive to issue them.

Discharge flow-reduction practices 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

Table 6. Summary of the nonparametric upper prediction levels on contaminant loading data.

	Did Not Exceed		Did Exceed	
Significance Level	5%	1%	5%	1%
The Number of Discharges Having Two Successive Post-Remining Samples Exceeding the Pre-Remining Maximum				
Acidity	39	26	8	5
Iron	42	28	5	3
Sulfate	29	21	13	5
Flow	38	26	9	5
The Number of Discharges Having Three Successive Post-Remining Samples Exceeding the Pre-Remining Maximum				
Acidity	47	42	5	5
Iron	47	42	5	5
Sulfate	37	32	10	10
Flow	49	44	3	3

Table 7. Summary of analyses of acidity load before and after remining. N indicates treatment was not initiated; under the PA system type 1, the post-remining median was not significantly different from pre-remining median at the 95% confidence level; under the PA 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 2 or 3 consecutive samples exceeding the pre-remining maximum. Y indicates that treatment was initiated under the given system. X indicates data were not available or were insufficient to complete the analyses.

	PA System		Mann-Whitney	Gibbons NUPL Method			
				2 Sample		3 Sample	
Site	Type 1	Type 2	U Test	1%	5%	1%	5%
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	X	Y	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	X	Y	Y	Y
9	N	N	N	X	X	X	N
10	N	N	N	N	N	N	N
11	N	N	Y	X	N	X	N
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	X	N	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	X	N	N	N
21	N	N	N	N	N	N	N
22	N	N	N	N	N	N	N
23	N	N	Y	X	N	N	N
24	N	N	N	N	N	N	N

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 amounts 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 flow-reduction option because it is the least expensive method of reducing vertical recharge. Furthermore, it must be performed to satisfy the reclamation requirements.

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

Although this study was conducted exclusively in Pennsylvania, similar remining programs in other Appalachian states should be at least as successful in terms of pollution load reduction. The geologic and hydrologic conditions in these other states do not differ that substantially from those of the western Pennsylvania coal fields. In order to qualify for a remining permit in Pennsylvania, an operator must demonstrate that the mining and instituted abatement practices have the potential to improve the water quality. Other states would need to apply a similar restriction during the permitting process.

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

REFERENCES

Ackerman, J.P, P.S. Campion, and E.B. Persson, (Preliminary) Final Report Deer Park Daylighting Project. U.S. E.P.A.

Berryman, D., B. Bobee, D. Cluis, and J. Haemmerli, 1988. Nonparametric Tests for Trend Detection in Water Quality Time Series. Water Resources Bulletin 24(3):545-556.

Davis, J.C. 1986. Statistics and Data Analysis in Geology, John Wiley & Son, Inc., New York, New York.

Gibbons, R.D, 1990. A General Statistical Procedure for Ground-Water Detection Monitoring at Waste Disposal Facilities. Ground Water 28(2):235-243.

Harris, J., J.C. Loftis, and R.H. Montgomery, 1987. Statistical Methods of Characterizing Ground-Water Quality. Ground Water 25(2):185-193.

Hawkins, J.W., 1993. Statistical Characteristics of Coal Mine Discharges in Western Pennsylvania, In-Press.

Helsel, D.R. 1983. Mine Drainage and Rock Type Influences on Eastern Ohio Stream Water Quality. Water Resource Bulletin. v. 19, no. 6. pp 881-887.

Helsel, D.R. 1987. Advantages of Nonparametric Procedures for Analysis of Water Quality Data. Hydrological Sciences Journal. v. 32, no. 2. 179-190.

Hirsch, R.M. and J.R. Slack, 1984. A Nonparametric Trend Test for Seasonal Data with Serial Dependence. Water Resources Research 20(6):727-732.

Lindorff, D.E., 1980. Hydrogeology of Surface Coal Mines in Illinois. Symposium on Surface Mining, Sedimentology and Reclamation. Lexington, KY:33-40.

Lineberry, G.T. and K.F. Unrug, 1990. Estimating Secondary Mining Potential of Inactive and Abandoned Appalachian Highwalls. International Journal of Surface Mining and Reclamation 4:11-19.

Montgomery, R.H., J.C. Loftis, and J. Harris, 1987. Statistical Characteristics of Ground-Water Quality Variables. Ground Water 25(2):176-184.

O'Steen, W.N. and H.W. Rauch, 1983. Effects of Coal Mining on Ground-Water Quality in West Virginia. 1983 Symposium on Surface Mining, Sedimentology and Reclamation. Lexington, KY:355-363.
Pennsylvania Department of Environmental Resources, The Pennsylvania State University, Kohlmann Ruggiero Engineers, 1988. REMINE.
Razem, A.C., 1983. Ground-Water Hydrology Before During, and After Coal Strip Mining of a Small Watershed in Coshocton County, Ohio. U.S. Geological Survey, Water-Resources Investigations Report 83-4155.
Reed, L.A., 1980. Effects of Strip Mining the Abandoned Anna S Mine on the Hydrology of Babb Creek, Tioga County, Pennsylvania. U.S. Geological Survey, Water Resources Investigations 80-53.
Richardson, A.R. and M. T. Dougherty, 1976. Feasibility Study Deer Park Daylighting Project. U.S. E.P.A. EPA-600/2-76-110.
Smith, M.W. 1988. Establishing Baseline Pollution Load from Preexisting Pollutional Discharges for Remining in Pennsylvania. In Proceedings: Mine Drainage and Surface Mine Reclamation Vol. II: Mine Reclamation, Abandoned Mine Lands and Policy Issues. U.S. Bureau of Mines IC 9184. Pittsburgh, PA. pp. 311-318.
Snedecor, G.W. and W.G. Cochran. 1971. Statistical Methods. The Iowa State University Press, Ames Iowa. 593 p.
Tukey, J.W., 1976. Exploratory Data Analysis. Addison-Wesley Publishing Co. Reading, MA.
Veil, J.A., 1993. Coal Remining: Overview and Analysis, US E.P.A. Contract W-31-109-Eng-38, 38 p.
Velleman, P.F. and D.C. Hoaglin, 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Wadsworth, Inc., Belmont CA.