Proceedings America Society of Mining and Reclamation, 2004

GEOCHEMICAL CLUSTER ANALYSIS OF MINE WATER QUALITY WITHIN THE MONONGAHELA BASIN[1]

James M. Stiles,[2] Joseph Donovan, David A. Dzombak, Rosemary C. Capo, and
Larry Cook

Abstract. The objective of this research was to conduct a cluster analysis of mean water quality data of 1,624 samples collected from 84 mine discharge sites within the Monongahela River basin over a 10 year period. This analysis produced four basic clusters. These four clusters were distinguished primarily by three factors: total dissolved solids, degree of acid neutralization, and mine discharge maturity. Most of the mine discharge sites (84%) were classified into a single cluster by the level one cluster analysis. Most of the discharges in this dominant grouping were older discharges from mines abandoned more than 15 years prior to discharge sampling. The discharges in the other three level one clusters exhibiting higher TDS levels tended to be more acidic and also to be from active mines, mines in the process of flooding, or mines flooded for less than 15 years

Future research should be devoted towards identifying the discharge maturity and flooding status of the various mine sites within the dataset, as well as the coal seam and overburden characteristics. The identification of these characteristics for all of the mine discharge sites would permit the analysis of separate datasets with comparable characteristics.

---

[1] Paper was presented at the 2004 National Meeting of the American Society of Mining and Reclamation and The 25th West Virginia Surface Mine Drainage Task Force, April 18-24, 2004. Published by ASMR, 3134 Montavesta Rd., Lexington, KY 40502.

[2] James M. Stiles is Environmental Engineer, West Virginia Water Research Institute, West Virginia University, Morgantown, WV 26506-6064, jim.stiles@mail.wvu.edu. Joseph J. Donovan is Associate Professor of Geology, Department of Geology and Geography, West Virginia University, Morgantown, WV 26506-6300. David A. Dzombak is Professor of Civil Engineering, Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890. Rosemary Capo is Associate Professor of Geology, Department of Geology and Planetary Science, 200 SRCC Building, 4107 O'Hara Street, University of Pittsburgh, Pittsburgh, PA 15260-3332. Larry Cook is a Graduate Student in the Department of Geology and Geography, West Virginia University, Morgantown, WV 26506-6300.

Proceedings America Society of Mining and Reclamation, 2004 pp 1819-1830
DOI: 10.21000/JASMR04011819
https://doi.org/10.21000/JASMR04011819

Introduction

Coal mining has taken place in the Monongahela Basin since the late eighteenth century, in a number of coal seams but especially the 6-8 ft. thick Pittsburgh coal seam. Since the beginning of mining operations, acid mine drainage (AMD) has been a problem in the Monongahela Basin. The goal of the research was to classify the 84 sampled mine discharges into empirical water quality groups, to assess the range of water types in the basin that need to be addressed in any basin wide AMD management planning. A secondary objective was to develop hypotheses about geochemical processes responsible for the observed water quality variability. The initial work plan was to employ a principal component analysis for statistical analysis; however, because the water chemistry data were in a number of constituents non-normal in distribution, the altered to include cluster analysis.

Methodology

Experimental work for this task consisted of preparing a consolidated database of water quality data for 1,624 samples collected by Carnegie Mellon University, University of Pittsburgh and West Virginia University researchers at 84 mine discharges sites within the Monongahela Basin. Because most of the data in the consolidated mine discharge database were obtained from sites at which multiple samples were collected over a ten year period, the cluster analysis was performed using mean water quality data for each site. The following water quality analyses were measured in the suite of samples: pH, total alkalinity, SO_4, Cl, Na, Ca, Mg, Fe, Mn, and Al. The mean values for these analyses are presented in Vandivort and Ziemkiewicz (2003). Most but not all of the analytical data reported were for dissolved concentrations measured on filtered samples. Analyses on unfiltered samples were assumed to represent primarily dissolved concentrations for the purposes of this analysis. The geographic locations of the sampled mine sites are shown in Fig. 1.

Normal concentrations, in units of milliequivalents per liter (meq/L) were calculated and used for the data analyses performed in this study. The normal concentrations of Mn and Fe were both calculated under the assumption that these metals were present exclusively in the divalent form at the time of sampling. Equivalent values of total alkalinity were calculated using the following identity: 1 meq/L of carbonate alkalinity is equal to 50 mg/L $CaCO_3$ carbonate alkalinity. The cumulative sum of equivalent concentrations for all ions was also calculated and indicated as “Sum” in the tables. Net alkalinity as used in this paper, is the total dissolved alkalinity as determined by titration, minus the sum of potential acidity from hydrolysis of Al^{3+}, Fe^{2+}, and Mn^{2+}, minus the actual hydrogen-ion acidity, and minus the acidity from the conversion of bisulfate to sulfate. This definition is effectively the converse of net acidity as defined by Hedin (2004).

Net alkalinity describes the acid-base balance between all dissolved acid-producing and acid-consuming solutes in the water, similar to the net neutralization potential (NNP) described by Joseph et al., (1994). If positive, this value would indicate that, after complete oxidation and hydrolysis of metals, there would be some residual alkalinity in the water and negligible metals. If negative, it would indicate that virtually all alkalinity would be consumed by hydrogen ion acidity generated by metal hydrolysis, and that supplemental base addition would be required to reduce the metals concentrations to low values. Net alkalinity values were calculated by subtracting the metal acidity from the total alkalinity in equivalent concentrations.

Cluster analysis is the clustering of similar objects within a dataset based on some quantitative measure of similarity. Cluster analysis has found wide application in applied scientific research and works well with data sets that contain outliers and other non-normal components. This robust character, however, comes at the expense of mathematical rigor (Hartigan, 1975).

For this analysis, the Ward minimum variance linkage method was employed as the similarity measure to distinguish the clusters, and the squared Euclidean distance method was used to determine the distances between the centroids of these clusters. The latter method was specifically recommended for use with the Ward linkage method by Lance and Williams (1967). Before the analysis was performed, all of the chemistry fields in the database were normalized to eliminate spurious scaling effects caused by different reporting units for the various constituents.

Cluster analysis sequentially groups observations into clusters, combining two un-grouped observations and/or clusters at a time according to the chosen objective function. Initially all the observations are un-clustered; finally, the entire dataset is clustered into a single group. To identify useful empirical classification of mine discharge chemistries, the intermediate-stage clusters were examined at points when there were two, three, four, or more clusters. The objective was to delineate natural groupings with low-variance clusters.

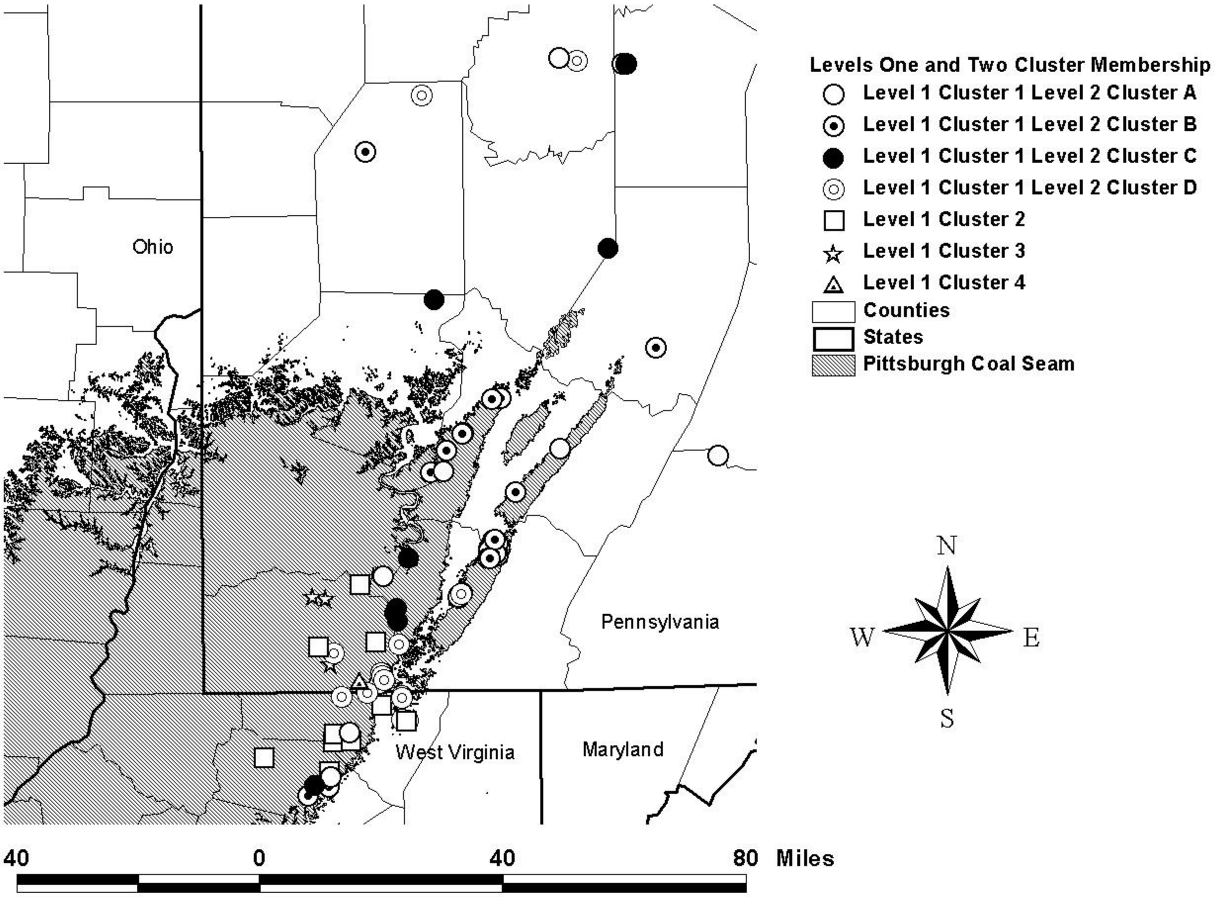

Figure 1. Geographic locations and cluster membership of the studied mine discharge sites.

The clustering procedure is illustrated by the dendrograph in Figure 2. The vertical axis in the dendrograph represents the squared Euclidean distance between the observations and clusters being combined by the horizontal lines in the figure. The vertical lines in Fig. 2 represent observations or clusters of observations. Vertical lines on the horizontal axis of the dendrograph represent the un-clustered observations at the beginning of the clustering process.

Results

Four clusters were chosen because of the relatively high value for the successive differences in the semi-partial R^2 values. The fourth hierarchical set of clusters – that with four groups of similar mine water chemistries – is the one chosen to have the most coherent classification of the mine waters. These classifications are given in Table 1 and are also shown in Fig. 1.

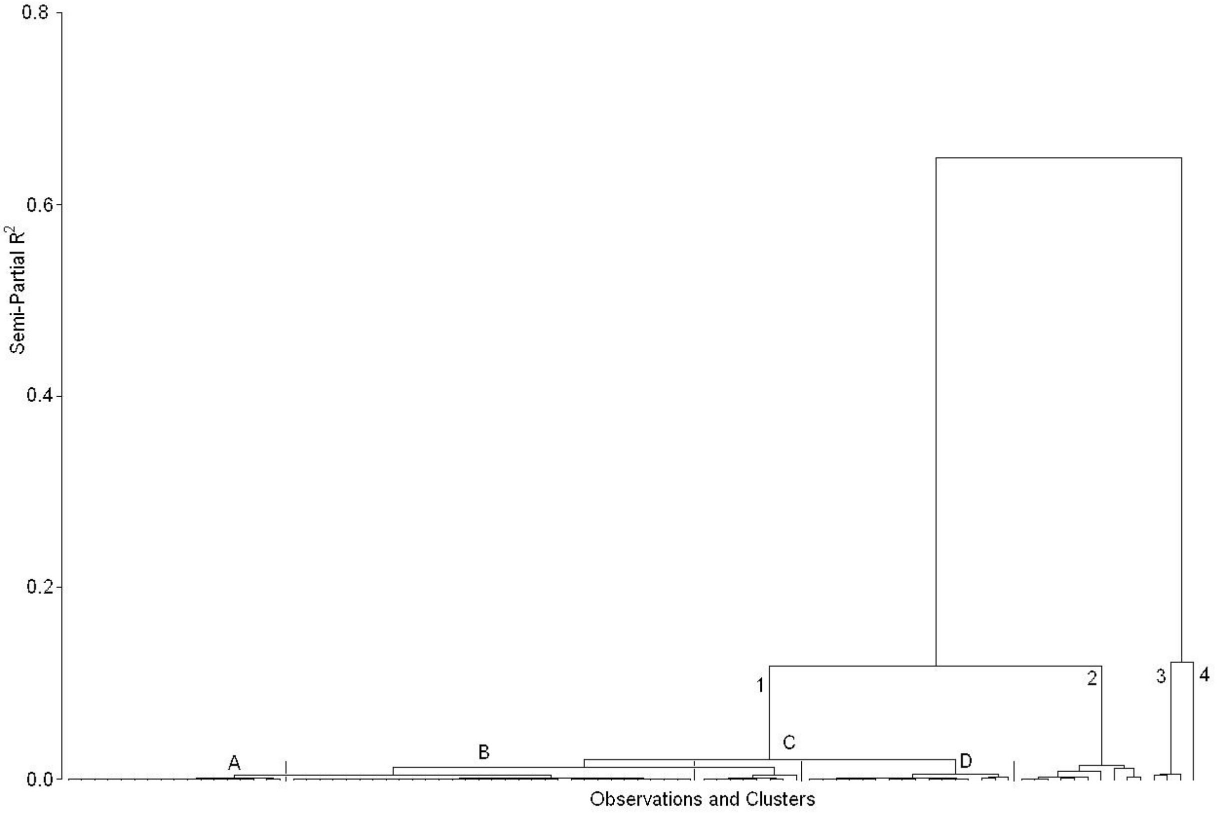

Figure 2. Dendrograph of the cluster analysis with the level one and two clusters shown.

Level One Analysis: Four Clusters

Tables 2 and 3 describe the chemistry of each cluster of mine discharge sites. Net alkalinity in Table 2 is defined as the total alkalinity minus the metal acidity in equivalent concentrations.

Table 3 lists these equivalent concentrations expressed as percentage of total cation or anion concentrations. For simplicity, the percentages are calculated based only on the ions shown, neglecting ions not included in the analysis and neglecting analytical error in charge balance. The non-inclusion of other ions induces relatively minor error. Thus all the cations in Table 3 add up to 100%, as do the anions, for each cluster.

Table 1. Level one cluster classification of sites.

Cluster	Sites
1	A0, H0, R3, S1, SANF, TAYL2A, TAYL2B, SEARS, MAXLD, HOWEB, LOWB, BIRD, MORR, LOWBM, STVIN, REDST, P0, B1, B2, H2, H4, M59, RD, RU, U1, U2, US60, PENNO, NORW, JENN, FILSON, EXPOR, DLM1, DLM2, IRW, COALR, GUFFU, GUFFL, DOUGR, WATER, IRWU, DLMLG, EXPART, WILS, A6, A7, AC, P1, BARR, GATES, PALMER, ADAH, BROWNS, PENNA, KEYST, R1, S03A, S06A, S13A, S14A, MAID3A, MAID3B, FETTY, BALLP, BOWL, TAYL4, CUMB2A, GRAYS, DESAL2, STRATT, R2
2	S22A, S24A, PAWPAW, HAGA, DOGL, LLEW, COLV, CUMB3, CLYDE
3	GRIME, RUFF2, CUMB1A
4	SHANN

Table 2. Median concentrations for the level one clusters.

			Dissolved Concentrations, meq/L								Calculated, meq/L	
Cluster	pH	Tot. Alk.	Na	Ca	Mg	Fe	Mn	Al	SO_4	Cl	Sum	Net Alk.
1	6.00	3.06	3.72	7.73	4.39	1.00	0.07	0.03	15.7	0.55	36.3	1.96
2	6.28	9.36	3.62	18.7	9.2	4.98	0.16	0.04	76.3	9.73	132	4.19
3	5.21	0.16	198	21.2	32.4	32.1	0.25	0.67	198	47.1	530	-32.9
4	4.49	0.47	156	18.1	62.4	129	1.22	9.10	421	5.67	803	-139

Table 3. Dominant ions in the level one clusters.

	--------Category--------			----Dominance----		Percentage of Total Cations or Anions, meq/L							
Cluster	pH	TDS	Net Alk.	Primary	Secondary	Na	Ca	Mg	Fe	Mn	Al	SO_4	Cl
1	Inter	Low	Sl. Alk.	SO_4, Ca	Mg, Na, Alk	22	46	26	6	0	0	81	3
2	High	High	Alkaline	SO_4, Ca	Mg	10	51	25	14	0	0	80	10
3	Inter	High	Acidic	SO_4, Na	Cl	70	7	11	11	0	0	81	19
4	Low	High	Acidic	SO_4, Na	Mg	42	5	17	34	0	2	99	1

Table 3 also classifies each cluster by pH as low (below 4.5), intermediate (4.5 - 6.0), or high (above 6.0), by TDS as high (above 100 meq/L) or low (below 100 meq/L), and by primary and secondary chemical constituent dominance. The primary constituents have more than 30% of the total anion or cation equivalent concentrations, while secondary ions have between 15% and 30% of the total anion or cation equivalent concentration. The interpretations of the clusters are as follows:

Cluster 1. Cluster 1 is characterized by variable levels of pH, alkalinity, calcium, aluminum, and chloride; and low levels of sodium, magnesium, iron, manganese, and sulfate.

Cluster 2. Cluster 2 is characterized by variable levels of pH, alkalinity, sodium, calcium, magnesium, manganese, aluminum, and chloride; moderate levels of sulfate; and low levels of iron.

Cluster3. Cluster 3 is characterized by low levels of pH and alkalinity; moderate levels of manganese and aluminum; and high levels of sodium, calcium, magnesium, iron, sulfate, and chloride.

Cluster4. Cluster 4 is characterized by very low levels of pH and alkalinity; moderate levels of chloride; high levels of sodium, calcium, manganese, and aluminum; and very high levels of magnesium, iron, and sulfate.

Level Two Analysis: Four Clusters within Level One Cluster 1

Because the majority (84%) of the studied mine sites were placed into cluster 1 by the level one cluster analysis, it was decided to examine this cluster more carefully with another level of analysis. This suggests that the Level One clustering was to a certain degree categorization of outliers sufficiently different from a single main group to resist integration with it.

Table 4 lists all of the sites within level one cluster 1 along with the corresponding level two cluster, which are also shown in Fig. 1. Tables 5 and 6 describe the chemistry of each level two cluster. Table 5 lists the median equivalent concentrations of the level two clusters. Table 6 lists the equivalent concentrations of the clusters expressed as percentage of total cation or anion concentrations. Like Table 3, the percentages are calculated based only on the ions shown, neglecting ions not included in the analysis and neglecting analytical error in charge balance. The interpretation of the four level two clusters is as follows:

Level Two Cluster A: SO_4-Ca (Mg-Na-Alk); Alkaline mine drainage. Cluster A is moderate in TDS and intermediate in pH. SO_4 is the dominant anion and Ca the dominant cation. All metals concentrations are low except Fe (2.6 meq/L).

Level Two Cluster B: SO_4-Ca (Mg-Na-Alk); Slightly alkaline mine drainage. Cluster B is virtually identical in distribution of ions to Cluster A, except it is lower in TDS and very slightly lower (0.8 versus 2.6 meq/L) in iron. It is also water dominated by SO_4-Ca with intermediate pH and moderate TDS. In terms of the Euclidean distance, these two clusters are the closer of the four level two clusters. All metals concentrations are low.

Level Two Cluster C: Alk-Na (SO_4-Ca); Peralkaline mine drainage. Cluster C is high in pH and moderate in TDS. It is a reversal of the trend for the other level two clusters, in that alkalinity and Na, instead of SO_4 and Ca, are the dominant ions. All metals concentrations are low.

Level Two Cluster D: SO_4-Mg (Ca-Na); Acidic mine drainage. Cluster D is the only level two cluster that is low in pH (median 4.33); it is also high in TDS. The distribution of cations is approximately equal between Mg, Ca, and Na.

Table 4. Level two cluster classification of sites.

Cluster	Sites
A	A0, H0, R3, S1, SANF, TAYL2A, TAYL2B, SEARS, MAXLD, HOWEB, LOWB, BIRD, MORR, LOWBM, STVIN, REDST, P0
B	B1, B2, H2, H4, M59, RD, RU, U1, U2, US60, PENNO, NORW, JENN, FILSON, EXPOR, DLM1, DLM2, IRW, COALR, GUFFU, GUFFL, DOUGR, WATER, IRWU, DLMLG, EXPART, WILS, A6, A7, AC, P1
C	BARR, GATES, PALMER, ADAH, BROWNS, PENNA, KEYST
D	R1, S03A, S06A, S13A, S14A, MAID3A, MAID3B, FETTY, BALLP, BOWL, TAYL4, CUMB2A, GRAYS, DESAL2, STRATT, R2

Table 5. Median concentrations for the level two clusters.

					Dissolved Concentrations, meq/L						Calculated	
Cluster	pH	Tot. Alk.	Na	Ca	Mg	Fe	Mn	Al	SO_4	Cl	Sum	Net Alk
A	5.64	5.90	6.04	11.33	6.30	2.61	0.26	0.63	22.42	0.66	56.16	2.39
B	5.25	2.41	2.78	6.17	3.32	0.76	0.11	0.51	11.31	0.85	28.22	1.03
C	6.93	24.77	13.57	4.58	1.88	0.74	0.09	0.08	5.13	1.59	52.42	23.85
D	4.33	1.95	10.78	12.94	14.64	2.97	1.01	5.99	36.63	5.93	92.83	-8.02

Table 6. Dominant ions in the level two clusters.

	----------Category----------			--------Dominance-------		Percentage of Total Cations or Anions, meq/L							
Cluster	pH	TDS	Net Alk.	Primary	Secondary	Na	Ca	Mg	Fe	Mn	Al	SO_4	Cl
A	Inter	Low	Alkaline	SO_4, Ca	Mg, Na, Alk	22	42	23	10	1	2	77	2
B	Inter	Low	Sl. Alk.	SO_4, Ca	Mg, Na, Alk	20	45	24	6	1	4	78	6
C	High	Low	Peralkaline	Alk, Na	SO_4, Ca	65	22	9	4	1	0	16	5
D	Low	High	Acidic	SO_4, Mg	Ca, Na	22	27	30	6	2	12	82	13

Discussions

The cluster analysis (CA) required two different levels of analysis. In the first level, about 84% of the observations clustering into a single group, and the rest into three much smaller "outlier" clusters. The smaller clusters are associated with waters that are high to very high in dissolved solids content, while the single large first cluster included more dilute waters. Therefore, TDS is a primary discriminating variable between observations.

Within the high TDS clusters (a total of 15 observations), discrimination occurs mainly by Fe concentration (Cluster 4 > Cluster 3 > Cluster 2) and alkalinity (Clusters 4 and 3 are low alkalinity and cluster 2 is high alkalinity). Clusters 3 and 4 are also elevated in Na and SO_4, the other dominant solutes in these mine waters. Clusters 3 and 4 consist of samples taken from a small number of mines that have yet to discharge water to the surface, and are either well samples taken from either flooding mines (Shannopin, Grimes, and Ruff Creek) or active mining operations (Cumberland Bleeder). Cluster 2 mines are from deep, generally large fully-flooded mines producing water of with high alkalinity and dissolved solids content.

Within the low TDS level one cluster (1), four additional groups may be resolved. The first two, A and B, are slightly to moderately net alkaline with intermediate levels of pH and TDS. They are very similar to each other, except for minor differences in Fe and TDS concentrations, and are both SO_4-Ca dominant. Level two cluster C is peralkaline mine drainage and quite different than the other clusters in that the dominant ions are alkalinity and Na with a high pH. Level two cluster D is acidic mine drainage that is dominated by SO_4, Mg, Ca, and Na ions, but the TDS concentration is not sufficiently high to include it with level one clusters 2, 3, and 4.

In summary, the cluster analysis suggests that TDS and extent of neutralization are important elements of the water chemistry. There are both low TDS and high TDS clusters within the dataset, and which can be further divided into acidic and alkaline sub-clusters. We infer that a principal cofactor favoring high TDS in these waters is the age, or maturity, of specific mine waters. Virtually all of the level one cluster 2, 3, and 4 waters are from mines that are either flooding, active, or flooded for less than 15 years. While the precision of knowledge regarding duration of flooding is poor, it can be safely said that only flooding mines populate level one clusters 3 and 4, and recently flooded mines dominate cluster 2. Therefore, these are immature mine waters.

We also believe that level one cluster 1, which contains samples taken from mature, low TDS waters, may be subdivided by this classification into three principal categories: alkaline (A and B); acidic (D); and peralkaline (C). There is a pronounced difference in chemistry between the peralkaline cluster and all the other level two clusters, resulting in dominance of alkalinity and Na ions rather than a dominance of SO_4 and either Ca or Mg ions, the common signature for the other clusters.

Level two cluster C contains a small number of observations with a small (less than 104 mg/L) Cl concentration; therefore, it is interpreted that the Na is possibly derived from an ion exchange mechanism, involving the loss of calcium and corresponding increase in alkalinity while maintaining calcite equilibrium similar to what was described by Capo et al., (2001) and Winters et al., (1999). All but one of the peralkaline waters are from a deep pool of the Pittsburgh coal mines on the east side of the Monongahela River, where the mines have been closed for over 20 years. The median alkalinity for this cluster exceeds 24.7 meq/L (1238 mg/L as $CaCO_3$) and is substantially higher than any other mine waters observed to date in the basin. The cofactors favoring variations in neutralization behavior, as defined by the level two clusters, are thought to be more complex, and would require additional work to resolve.

Conclusions

The results of the level one cluster analysis indicate that the quality of any Monongahela Basin mine discharge can be classified into one of four basic groups according to the dissolved solids content, with most of the sites (84%) falling into the category of low TDS waters exhibiting variable pH. Within this dominant group of discharges identified by the level one cluster analysis, a level two analysis indicated that a mine discharge could be placed into one of three principal categories: alkaline, acidic, or peralkaline. Mine discharge sites within the alkaline and acidic categories were dominated by SO_4 and either Ca or Mg ions and distinguished by net alkalinity. Mine discharge sites within the peralkaline category were dominated by alkalinity and Na ions. Most of the discharges in this dominant grouping were older discharges from mines abandoned more than 15 years prior to discharge sampling. The discharges in the other three level one clusters exhibiting higher TDS levels tended to be more acidic and also to be from active mines, mines in the process of flooding, or mines flooded for less than 15 years.

Future research should be devoted towards identifying the relative maturity of specific discharges and flooding status of the individual mine sites from which they flow. In addition, those samples collected by pumping from mines which had not yet completely flooded and have yet to discharge to the surface may have distinctive chemical signatures and should be analyzed separately from the waters of "mature discharges", which have been taking place for long periods. The results of the current analysis should also be viewed in light of the coal seam chemistry, overburden characteristics, and flooding status to further account for the differences in water chemistry between the various mine discharge sites.

Acknowledgments

This article was prepared with the support of the U.S. Department of Energy - National Energy Technology Laboratory and Parsons, Inc., under Contract Agreement No. DE-FC26-98FT40396 and the DOE-NETL University Partnership. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the Department of Energy - National Energy Technology Laboratory and Parsons, Inc.

Literature Cited

Capo, R.C., Winters, W.R., Weaver, T.J., Stafford, S.L., Hedin, R.S., Stewart, B.W., 2001. "Hydrogeologic and geochemical evolution of deep mine discharges, Irwin syncline, Pennsylvania," Proceedings of the Twenty-Second West Virginia Surface Mine Drainage Task Force Symposium, p 144-153.

Hartigan, J.A. (1975), Clustering Algorithms, John Wiley and Sons Inc., New York, NY.

Hedin, R.S. (2004), “The use of Measured and Calculated Acidity Values to Improve the Quality of Mine Drainage Datasets,” this volume.

Joseph, W.L., Brady, K.B.C., and Perry, E.F., 1994. “Variation in Computing Acid Base Accounting in the U.S.” Abstract In: International Land Reclamation and Mine Drainage Conference and Third International Conference on the Abatement of Acidic Drainage, U.S. Bureau of Mines, Special Publication SP 06B-94, p 417. `This is an abstract`

Lance, G.N. and Williams, W.T. (1967), "A General Theory of Classificatory Sorting Strategies, I. Hierarchal Systems," Computer Journal, 9, pp 373-380. http://dx.doi.org/10.1093/comjnl/9.4.373.

Vandivort, T., Ziemkiewicz, P., 2003. WV173 Phase III EPA Region III Mine Pool Project Final Report. Contract Report to Parsons Inc. under U.S. Department of Energy Cooperative Agreement Number DE-AM26-99FT40463.

Winters, W.R., Capo, R.C., Wolinsky, M.W., Weaver, T.J., and Hedin, R.S., 1999. “Geochemical and hydrogeologic evolution of alkaline discharges from abandoned mines,” Proceedings from the Sixteenth Annual International Pittsburgh Coal Conference, October 11-15, 1999, CD-ROM, ISBN 1-890977-16-0, section 6-5, p 1-36.