VARIABILITY OF PHYSICOCHEMICAL ANALYTICAL RESULTS FROM MINESOILS AND QA/QC CONSIDERATIONS[1]

by
Jon E. Brandt and Jan K. Horbaczewski[2]

Abstract. The Texas Mining and Reclamation Association (TMRA) has sponsored a soil sample round robin program since 1990. To date, 17 different soil samples (including blind duplicate samples M and P) have been analyzed in a series of six rounds of analysis. Five laboratories have participated in this program. Three of these were commercial laboratories, one was an electric utility's in-house laboratory, and one was the state regulatory authority's laboratory. Samples submitted for analysis included minesoils and native soils. Results indicate that average inter-laboratory variability was approximately five times greater than intra-laboratory variability. The level of variation was affected by the sample being analyzed and the analytical parameter that was involved. The larger inter-laboratory variability suggest that a greater convergence of analytical results may be attained through a rigorous examination of individual laboratory methodologies and implementation of widely-used standard operating procedures (SOPs). An examination of coefficient of variation (C.V.) data from twelve analytical parameters indicated that pH was the least variable and trace elements were among the most variable. The inclusion of near minimum detection level (MDL) values tended to increase the C.V. of data. Analyses of variance (ANOVA) indicated that the sources of variation of most parameters were inconsistent over time. Differences in equipment, sample extraction, and personnel could also account for different values among the laboratories. Analytical parameters that exhibited a sample by laboratory (S X L) interaction suggest that there was inconsistency among laboratories over time. Samples M and P were blind duplicates of the same sample, included in the round robin to assess the precision of the participating laboratories. Results of samples M and P indicate that there was either a significant laboratory main-effect or S X L interaction for most of the analytical parameters, suggesting that either lab inconsistencies or sample heterogeneity could be the cause. The use of a standard reference material (SRM), with known parameter values, can help identify which laboratories are accurate. There are analytical considerations that require attention, such as sample preparation and pretreatment, sample extraction, and both determinate and indeterminate errors. Sample preparation is a critical step where variability can be introduced. The development and implementation of SOPs are essential steps for any laboratory. The variable results obtained by the TMRA round robin program illustrate the need for QA/QC procedures and the use of good laboratory practices. The program's results also prompted the following questions: 1) are the levels of variability of sufficient importance to warrant additional research and 2) are there practical mine reclamation implications to the variability exhibited in the round robin data?

Additional Key Words: laboratory analysis, inter-laboratory variability, minesoil analysis, geochemistry, standard reference material (SRM), Texas

Introduction

The purpose of this paper is to evaluate the results of the soil laboratory round robin program sponsored by the Texas Mining and Reclamation Association (TMRA). This paper examines some of the data that have been obtained and relates this information to basic QA/QC (quality assurance and quality control) principles. It will look at variability among laboratories (inter-laboratory error) and within individual laboratories (intra-laboratory error) and will document the observed variability of a limited number of analytical parameters.

---

[1] Paper presented at the 1997 National Meeting of the American Society for Surface Mining and Reclamation, Austin, Texas, May 10-15, 1997.

[2] Jon E. Brandt is Soil Scientist, Railroad Commission of Texas, Surface Mining and Reclamation, P.O. Box 12967, Austin, Texas 78711-2967; Jan K. Horbaczewski is Environmental Manager, Morrison Knudsen Corporation, P.O. Box 400090, San Antonio, Texas 78229.

Proceedings America Society of Mining and Reclamation, 1997 pp 152-166
DOI: 10.21000/JASMR97010152

https://doi.org/10.21000/JASMR98010152

The TMRA round robin program was initiated to identify potential problems caused by discrepancies in results among laboratories performing soil, overburden and spoil analyses for the Texas coal mining industry and to offer possible recommendations to correct any problems these differences may cause.

A sampling program should strive to provide samples with as little total variance as possible and with the best available accuracy and precision. Analytical precision is generally measured in the vicinity of 1%, while sample variation in soils can range up to 30% and above (Barth and Mason, 1984). The TMRA round robin program focused on the analytical portion of the sampling program, attempting to measure the sources of variability after a sample is collected and prepared. Variability in soil and minesoil data can come from a number of sources: sample location (spatial variability), time of sampling (temporal variability), sample handling and preparation; sample aliquot selection; analytical methodology (extraction/dissolution methods); instrumentation; gross errors (operator mistakes, transcription errors); and random error. Sample collection procedures provide the main limitations on data precision (Munk et al., 1996).

The TMRA round robin program began in 1990 with the initial participation of four laboratories, which then increased to five with Round 4. Three of the laboratories were commercial, one was an in-house electric utility laboratory, and one was a part of the state regulatory authority. As of January 1997, six rounds of analyses have been completed (three samples per round), at a rate of approximately one round per year (Table 1).

This paper examines the variability exhibited by twelve geochemical parameters that are generally of most concern to the Texas mining industry: pH; sand and clay contents; electrical conductivity (E.C.); cation exchange capacity (CEC); neutralization potential; total sulfur; pyritic sulfur (potential acidity); exchangeable acidity; acid/base account (ABA); total cadmium; and total selenium. However, the TMRA round robin program encompassed a much broader span of over 45 different geochemical measures, including major anions (chloride, carbonate, bicarbonate, and sulfate); major soluble cations (sodium, potassium, calcium, and magnesium); major exchangeable cations (sodium, potassium, calcium, and magnesium); all texture parameters (sand, silt, and clay); all sulfur forms (total, pyritic, sulfate, and residual/organic); organic carbon; exchangeable aluminum; base saturation; sodium adsorption ratio; arsenic; boron; chromium; copper; manganese; molybdenum; nickel; lead; uranium; and zinc. This paper will attempt to accomplish five objectives: 1) compare inter-laboratory to intra-laboratory variability; 2) examine the effect of data values on the calculated variability of analytical parameters, since values near the minimum detection levels may appear to exaggerate the amount of variability; 3) identify some of the potential sources of variability in analytical data; 4) identify procedural factors that may affect analytical results; and 5) re-examine several QA/QC principles that must be considered in minesoil analyses.

Table 1. History of TMRA round robin program.

Round	Sample I.D.	Number of Repli-cations	Number of Parti-cipating Labs	Date of Final Report
1	A, B, C	0	4	Sept. 28, 1990
2	D, E, F	2	4	July 5, 1991
3	G, H, I	2	4	Sept. 21, 1992
4	J, K, L	2	5	Nov. 8, 1993
5	M, N, O	2	5	July 8, 1994
6	P, Q, R	2	4	July 31, 1995

Materials and Methods

The samples used in the round robin program were collected from different Texas mine sites in different geochemical conditions, to provide as wide a range of properties as possible. It should be noted that Sample P was a blind duplicate (i.e., unknown to the laboratories) of Sample M.

The following procedures were used to prepare the samples for use in the round robin program: 1) field sample was collected in a 5-gallon pail and brought to one laboratory; 2) field sample was air-dried and lightly ground; 3) field sample was passed through 2-millimeter sieve; 4) fine earth fraction (material passing through 2-millimeter sieve) was subdivided into smaller portions (of about 2 kg) using a riffler; 5) sub-samples (approximately 2 kg) of the fine earth fraction were sent to each laboratory participating in the program; 6) each laboratory prepared its own duplicates from the 2 kg sub-sample; and 7) each laboratory performed analyses in parallel on its two duplicate subsamples.

The laboratory methods used for the analysis of native soil, overburden and post-mine soil materials were those recommended by Texas' coal mining regulatory authority, described in *Overburden Parameters and Procedures* (Railroad Commission of Texas, 1989) and detailed in Technical Release SA-2: *Materials Suitable for Placement in the Top Four Feet of Leveled Minespoil, Including Topsoil Substitutes* (Railroad Commission of Texas, 1988). These were the methods employed in TMRA's round robin program, unless otherwise indicated (Table 2). More recently (January 1, 1996), the Railroad Commission has issued Advisory Notice AG-RP-145 *Overburden and Minesoil Samples - Preparation Procedure for Large and Small Volume Samples*, but this has not been used to date in TMRA's round robin program.

The data were first examined by tabulating the ranges in parameter values obtained for each sample. The inter-laboratory range was determined by subtracting the minimum value from the maximum value for each of the parameters for every sample. The intra-laboratory range was estimated by subtracting the minimum difference between replicates observed from any of the laboratories from the maximum difference between replicates observed from any of the other laboratories (for each sample and analytical parameter).

A multiple analysis of variance (ANOVA) was then performed for each analytical parameter, employing the following main effects (the selected sources of variability for the statistical model): sample, laboratory, and replication (Type III sum of squares; Manugistics, 1992). A limited number of samples and laboratories were used for the statistical analyses because not all of the laboratories participated in each round of samples (the ANOVA employed data from laboratories A, C, D, and E and samples G, H, I, J, K, L, M, N, O, P, Q, and R). The laboratory types (government, commercial, etc.) will not be identified, since the laboratories were assured anonymity.

Results and Discussion

Inter-Laboratory versus Intra-Laboratory Variability

A preliminary comparison was made between the inter-laboratory and intra-laboratory variability, by examining the value ranges for the 12 analytical parameters (Table 3). The comparison shows that, on average, there was a 1.00 to 0.22 ratio between the inter- and intra-laboratory ranges; i.e., the variability among laboratories was approximately five times the difference between replicates performed within each laboratory. Seven out of twelve of the parameters had an average intra- to inter-laboratory (IIL) ratio between 0.20 to 0.27. Two analytical parameters had relatively high IIL ratios, near 0.10: electrical conductivity and exchangeable acidity.

Table 2. Analytical procedures used for theTMRA round robin program.

Analytical Procedure	Notes	Reference
pH	1:1 soil-water extract, pH measured after 1 hr.	Black, C.A., 1965
Texture (Particle Size Distribution)	Hydrometer determination, clay fraction measured at 12 hrs.	Sobek, et al., 1978
Electrical Conductivity (E.C.)	1:1 soil-water extract, extracted after 1 hr. equilibration time	Black, C.A., 1965
Neutralization Potential (NP)		Sobek et al., 1978
Total Sulfur		Sobek et al., 1978
Pyritic Sulfur		Sobek et al., 1978
Exchangeable Acidity (EA)	KCl-exchangeable acidity; run on samples below pH 5.5	Page, A.L., 1982
Acid/Base Account (ABA)	ABA = NP - (PA + EA)	Railroad Commission, 1988
Total Cadmium (Cd)		Bajo, 1978; EPA, 1979
Total Selenium (Se)		Bajo, 1978; EPA, 1979

Table 3. Comparison between inter- and intra-laboratory parameter value ranges (samples M and P were the same sample).

	Inter-Laboratory Ranges †												Intra-Laboratory Ranges											
Sample	pH	Sand	Clay	E.C.	C.E.C.	NP	T_S	Pyr.	EA	ABA	Cd	Se	pH	Sand	Clay	E.C.	C.E.C	NP	T_S	Pyr.	EA	ABA	Cd	Se
A	0.4	5	5	4.1	2	10	0.23	0.12	9.4	19.4	0.7	0.4	--	--	--	--	--	--	--	--	--	--	--	--
B	1.2	1	3	0.1	0	4.2	MDL	MDL	0.3	4.6	0.3	0.4	--	--	--	--	--	--	--	--	--	--	--	--
C	1.5	2	3	0.1	3	4.4	MDL	MDL	0.3	4.3	0.5	0.5	--	--	--	--	--	--	--	--	--	--	--	--
D	0.3	6	2	0.0	2	0.9	MDL	MDL	0.4	0.6	0.4	0.4	0.1	3	1	0.0	0	0.1	MDL	MDL	0.0	0.0	0.0	0.1
E	0.2	6	4	0.1	5	0.9	0.02	MDL	MDL	1.2	0.5	1.0	0.0	1	1	0.0	0	0.2	.01	MDL	MDL	0.2	0.1	0.1
F	0.2	9	3	0.5	4	7.5	0.31	0.72	MDL	10.9	1.1	0.6	0.0	1	1	0.0	4	1.1	0.15	0.20	MDL	1.3	0.1	0.2
G	0.5	2	7	0.1	1	1.9	MDL	MDL	0.8	2.1	0.4	0.9	0.2	1	2	0.0	1	1.2	MDL	MDL	0.05	0.4	0.0	0.9
H	0.3	6	14	0.1	1	1.9	MDL	MDL	MDL	1.9	0.4	0.1	0.1	2	1	0.1	0	0.3	MDL	MDL	MDL	0.2	0.0	0.1
I	0.2	7	7	1.4	1	1.5	0.05	0.05	1.2	4.5	0.5	1.4	0.1	2	0	0.1	1	0.3	0.03	0.02	0.2	0.6	0.1	0.1
J	0.3	13	9	0.4	14	6.2	0.38	0.12	MDL	3.7	0.4	2.0	0.0	1	4	0.1	2	1.9	0.25	0.01	MDL	0.3	0.1	0.1
K	0.9	7	5	0.1	11	3.8	0.03	MDL	15	3.7	0.0	1.1	0.1	0	1	0.0	1	1.0	0.01	MDL	0.6	0.1	0.0	0.0
L	0.8	14	5	1.7	33	3.1	0.04	0.07	MDL	12.5	0.2	0.2	0.2	4	3	0.2	5	0.3	0.01	0.02	MDL	0.2	0.0	0.0
M	0.5	9	6	0.2	5	1.3	MDL	MDL	MDL	3.6	0.4	1.0	0.2	4	1	0.1	0	0.2	MDL	MDL	MDL	0.2	0.0	0.0
N	0.4	8	6	0.1	16	4.0	MDL	MDL	1.4	1.3	0.8	1.0	0.2	2	2	0.0	2	2.4	MDL	MDL	0.2	0.8	0.0	0.5
O	0.3	9	6	1.2	--	1.5	0.73	0.44	2.8	0.8	0.6	1.0	0.0	1	2	0.2	--	0.4	0.16	0.06	0.8	4.7	0.0	0.0
P	0.4	15	8	0.3	4	1.3	MDL	MDL	MDL	1.5	0.3	1.3	0.1	4	4	0.1	1	0.2	MDL	MDL	MDL	0.2	0.3	0.2
Q	0.6	8	13	3.9	7	13.5	0.03	MDL	MDL	12.4	0.4	1.4	0.1	1	1	0.6	3	1.3	0.01	MDL	MDL	0.9	0.2	0.1
R	0.2	11	10	1.4	16	14.1	0.23	0.07	MDL	14.1	0.6	1.5	0.1	2	1	0.5	2	11.5	0.01	0.01	MDL	2.6	0.1	0.1
Mean	0.5	7.7	6.4	0.9	7.4	4.6	0.2	0.23	3.5	5.8	0.5	0.9	0.1	1.9	1.7	0.1	1.6	1.5	0.07	0.05	0.3	0.8	0.1	0.2
Ratio of Mean Intra-Lab Ranges to Mean Inter-Lab Ranges:													0.20	0.25	0.27	0.11	0.22	0.33	0.35	0.22	0.09	0.14	0.20	0.22

† Units for both the inter- and intra-laboratory ranges are as follows: pH, standard units; sand, %; clay, %; E.C. - electrical conductivity, mmhos/cm; C.E.C. - cation exchange capacity, meq/100g; N.P.- neutralization potential, tons/1000 tons; T_S-total sulfur, %; Pyr.-pyritic sulfur, %; EA-exchangeable acidity, meq/100g; ABA-acid/base account, tons/1000 tons; Cd-total Cd, ppm; and Se-total Se, ppm. MDL - minimum detection limit of the analysis.

These relatively high IIL ratios indicate that inter-laboratory variability for these parameters is approximately ten times as much as that observed in replicate samples run by the laboratories. A high IIL ratio (in relation to the other analytical parameters) may suggest that there are differences in the methodologies used by the laboratories. The larger inter-laboratory variability suggest that a greater convergence of analytical results may be attained through a rigorous examination of individual laboratory methodologies and development of widely-used standard operating procedures (SOPs).

Neutralization potential and total sulfur exhibited the lowest IIL ratios, at 0.33 and 0.35, respectively. These relatively low IIL ratios may suggest a couple of possibilities: 1) the variability of the results may be due to inherent procedure variability or 2) the variability of the results may be caused by the heterogenous nature of the materials being analyzed. Mixed overburden samples, sometimes comprised of several lithologic materials with varying levels of weathering, may exhibit greater variability of analytical results if the entire samples are not homogenized properly during sample preparation.

Coefficients of Variation - Analytical Methods

The variability of procedures is sometimes gauged by the coefficient of variation (C.V.) of the results. The C.V. is a measure of the spread of data and compares the amount of standard deviation (the positive square root of variance) relative to the mean value (C.V., % = Std. Dev./Mean X 100). However, relative variability between procedures may be exaggerated when sample values are near the minimum detection limit (MDL, the lowest level statistically different from a blank) of a particular analysis. A comparison between the results of two fictional samples, analyzed by five different laboratories, illustrates this phenomenon. The first sample has a mean total sulfur content of 0.07 ± 0.03%, resulting in a coefficient of variation (C.V.) equal to 43%. The second sample has a mean of 0.21 ± 0.03% total sulfur, resulting in a C.V. of 14%. The repeatability of values for both samples was the same (±0.03%), so there was likely no difference in the analytical procedures used for both samples, but the apparent spread of data in the first sample was three times the amount seen in the second sample.

There is an approach to evaluating procedure variability that takes into account the minimum detection levels (MDLs) of individual analyses and, consequently, minimizes the problems of exaggerated variability caused by low-value samples. A threshold value of 5 times the MDL has been suggested by the U.S. EPA (EPA, 1995) as a critical point when making statistical comparisons of analytical results, since values near the MDL may not be as reliable due to instrument drift and other errors. The C.V.s were calculated for the 12 analytical parameters examined in this paper. The C.V. ranged from 3.9% to 103% when all of the samples were included in the statistical calculations (Table 4). The C.V.s fell by an average of 45% of the original C.V. value when only samples above the threshold were considered. The number of samples used to calculate the above-threshold C.V.s also fell by 22 to 78 percent, thus showing the effect many of the samples had in increasing the apparent coefficients of variation. The C.V.s of the above-threshold results ranged from 3.9% (mean C.V. for pH) to 64% (mean C.V. for total Cd). Approximately half of the examined analytical parameters had C.V.s less than 20%.

Multiple-Factor Analysis of Variance

Two sets of analysis of variance (ANOVA) were performed. The first examined the data produced by four laboratories, analyzing 12 samples each. The second set of ANOVAs considered the results of analyses run on the same blind sample (samples M and P), thus attempting to measure the precision (repeatability) of four of the participating laboratories.

ANOVA - Four laboratories running eleven samples. A summary of the results of ANOVAs run on each parameter is presented in Table 5. Several important observations can be made with an initial examination of the statistical data: 1) there was no significant variability in results contributed by the replications performed by each laboratory; 2) there were no significant interactions between sample and either laboratory or replication (S X L or S X R); and 3) the samples provided a significant amount of variability for all of the parameters except total cadmium (since the levels of Cd observed in all of the samples were similar). A significant sample effect indicates that a significant source of data variability was due to differences in the samples analyzed, something that would be expected, since soil materials from very different environments had been selected. Many of the sample by laboratory (S X L) interactions were confounded; therefore, it was not possible to distinguish between the two sources of variability in those cases. The six analytical parameters where the S X L interaction was confounded, or where the labora-

Table 4. Comparison of coefficients of variation (C.V.s) for selected analytical parameters.

	All Values Included					Included Only Values Above Threshold				Included Only Values Below Threshold			
	Coefficient of Variation, % (C.V.)				Threshold	Coefficient of Variation, % (C.V.)				Coefficient of Variation, % (C.V.)			
Parameters	Mean	Sample No.	Min.	Max.	Value †	Mean	Sample No.	Min.	Max.	Mean	Sample No.	Min.	Max.
pH	3.9	18	1.2	12	None	3.9	18	1.2	12	N/A ‡	---	---	---
Sand, %	7.4	18	0.8	20	5% Sand	7.4	18	0.8	20	N/A	---	---	---
Clay, %	13	18	3.5	61	5% Clay	13	18	3.5	61	N/A	---	---	---
Elect. Cond., mmhos/cm	33	18	0	143	0.1 mmhos/cm	18	14	0	43	82	4	43	143
C.E.C., meq/100g	18	17	5.9	45	0.5 meq/100g	18	17	5.9	45	N/A	---	---	---
Neut. Pot., tkt	31	18	6.6	121	2 tkt	27	10	6.6	68	62	8	26	121
Total S, %	55	18	8.5	146	0.05% S	16	8	8.5	23	72	9	31	118
Pyritic S, %	71	18	0	250	0.05% S	26	4	11	41	90	14	0	250
Exch. Acid., meq/100g	46	18	0	158	0.5 meq/100g	18	8	6.4	42	69	10	0	158
Acid/Base Acct., tkt	59	18	6.2	167	5.5 tkt ¶	32	8	6.2	62	81	10	7	167
Total Cd, ppm	103	18	0	300	0.25 ppm	64	5	50	67	118	13	0	300
Total Se, ppm	76	18	29	200	0.25 ppm	59	8	29	80	89	10	30	200

† Threshold value is a product of the parameter's minimum detection level (MDL) times 5; threshold values have no relation to regulatory limits.

‡ Either no threshold value or none of the samples were below the corresponding threshold value.

¶ Value calculated as follows: Neut. Pot. (2 tkt) + Pyritic-S (0.05 X 31.25=1.6 tkt) + Total S (0.05 X 31.25=1.6 tkt) + Exch. Acid. (0.5 X .5 = 0.3 tkt); included pyritic-S and total-S used because both are used in calculating the potential acidity.

Table 5. Multiple factor analysis of variance for four laboratories, twelve samples, and two replications.

	Main Effects			Interactions		
Parameter	Sample (S)	Lab (L)	Replicate (R)	S X L	S X R	L X R
pH	*** †	***	NS	***	NS	NS
Sand, %	***	***	NS	***	NS	NS
Clay, % (12-hr reading)	***	***	NS	***	NS	NS
Elec. Cond., mmhos/cm	*	***	NS	***	NS	NS
C.E.C., meq/100g	***	***	NS	EC	NS	NS
Neut. Potential, tkt	***	**	NS	EC	NS	NS
Total Sulfur, %	***	NS	NS	EC	NS	NS
Pyritic Sulfur, %	***	*	NS	EC	NS	NS
Exch. Acidity, meq/100g	***	NS	NS	EC	NS	NS
Acid/Base Account, tkt	***	**	NS	EC	NS	NS
Total Cd, ppm	NS	***	NS	EC	NS	NS
Total Se, ppm	***	***	NS	EC	NS	NS

† *, **, *** significant at the 0.05, 0.01, and 0.001 probability levels, respectively; NS - not significant at the p = 0.05 level; EC - interaction between main effects was confounded.

tory was also a significant source of data variability, were CEC, NP, pyritic-S, ABA, total Cd, and total Se. These parameters will be briefly discussed below. The results of the significant main effects and their interactions will not be included in this paper. There were four parameters that had significant sample by laboratory (S X L) interactions: pH; sand and clay contents; and electrical conductivity. The significant S X L interactions indicate that there was inconsistent variability within the laboratories over time. There were no significant differences among laboratories for total sulfur or exchangeable acidity.

Cation Exchange Capacity (CEC). The CEC procedures of each laboratory should be compared to determine if slight differences in methodology may be producing the contrast in results.

Neutralization Potential (NP). The NP procedures used by each laboratory should be compared to determine if the differences in results are caused by variations in methodology. Incompletely homogenized sample materials may also cause data variability.

Pyritic Sulfur. Pyritic-S variability may have been caused by either differences in sulfur fractionation methodology, differences in instrumentation, or subsample variability introduced during the selection of sample aliquots.

The sulfur fractionation (separation of total sulfur into sulfate-S, pyritic-S, and organic/residual-S) procedures used by each laboratory should be examined carefully to see if differences in methodology may be contributing to variability of pyritic-S results. The pyritic-S value is critical to the determination of the ABA, since it indicates how much potential acidity (PA) may be generated. There is a lot of room for variation in potential acidity procedures, so it is imperative to know whether a direct determination method (hydrogen peroxide) or an indirect method (stoichiometric calculation using either pyritic-S or pyritic-Fe) is being used to estimate PA. Additionally, the indirect methods may use either sequential leachings performed on the same sample or separate leachings conducted on separate sample aliquots, which could result in different types of cumulative errors.

Acid/Base Account (ABA). Differences in ABA results are difficult to explain since ABA values are derived from the results of three separate analyses (NP, PA, and exchangeable acidity). Nevertheless, the procedures used by each laboratory should be examined carefully to see if differences in methodology may be contributing to ABA variability.

Total Cadmium and Selenium (Cd and Se). Differences in total metal values may be caused by variations in sample digestion (acid dissolution) and instrumentation (eg. atomic absorption spectroscopy versus inductively-coupled plasma atomic emission spectroscopy).

ANOVA - Same sample, run twice. A multiple -factor analysis of variance (ANOVA) was performed for the identical sample that was run on two separate occasions (identified as samples M and P). The analysis of this blind duplicate provided the opportunity to measure the precision (repeatability) of four of the laboratories. The results indicate that differences between same-lab replicates did not contribute a significant amount of variability to any of the results (Table 6). None of the sample or laboratory by replicate interactions (S X R or L X R) were significant. Several of the main effect interactions were either confounded or did not have enough degrees of freedom to perform the statistical analysis. The ANOVA for pyritic sulfur was not possible because most of the results were at or below the MDL. The ABA was the only parameter for which the sample effect was significant. This may have been due to the inherent variability of the sample; it could also have been caused by the cumulative variability of the three procedures involved. There were seven parameters for which the laboratory effect was significant, in addition to four parameters where sample by laboratory interactions (S X L) were significant.

Laboratories contributed a significant amount of variability for the sand and clay contents, electrical conductivity, cation exchange capacity, total sulfur, and total cadmium. Differences in sample texture, or particle size, results indicate there may be procedural differences between the laboratories (specifically with shaking/dispersion techniques). Laboratory differences in EC results were probably of little practical importance, since the results from all of the laboratories did not differ by more than 0.2 mmhos/cm. Slight differences in the procedures used by each laboratory, involving shaking and equilibration times, can cause the variation in EC results. Inconsistencies in the way MDL results are reported is probably the reason why there were differences in the total sulfur values (which ranged between 0 and 0.02%). Differences in total Cd could be due to differences in sample digestion procedures or instrumentation.

There were four parameters that had significant sample by laboratory (S X L) interactions: pH; neutralization potential; exchangeable acidity; and total Se. The significant S X L interactions indicate that there was inconsistent variability within the laboratories over time.

Analytical Methodology

The scope of this paper does not include a detailed examination of each of the analytical parameters. However, we felt it was worthwhile to mention several factors that may affect analytical results. Many of the factors involve common sense; but nevertheless, the round robin results indicate there are sources of variability that are not routinely identified and managed, and therefore deserve a brief review.

Sample Preparation and Pretreatment. Most soil laboratory analyses are performed on the fine earth fraction (material passing through a 2-mm aperture sieve). As a rule of thumb, finely ground materials tend to produce less variable data. It is necessary, then, to determine if increased grinding is compatible with the operational objectives of the analytical procedures.

The separation of the fine earth fraction from native soils is usually relatively easy to accomplish since the soils have a developed structure and disintegrate under light to moderate pressure. On the other hand, overburden materials and the post-mine soils from which they are derived (whether chemically oxidized or reduced) tend to retain the consolidation resulting from geological rock-forming processes. This is particularly true of reduced materials in which cementing agents have not been attacked by oxidation or other weathering processes. Thus, the silt and clay particles of these materials may be derived from rocks ranging from soft fissile shales to very indurated siltstones. The degree to which these rocks are ground will determine the kind of results that will be obtained. At one extreme, a hard siltstone could be ground to give a predominantly sandy/gravelly material; at the other extreme, the same rock could be milled to a silt. Recently, an attempt has been made by the RCT to standardize the procedure in Advisory Notice AG-RP-145, as indicated above. The authors' prior experience indicates that the sample preparation phase may introduce a much larger amount of variability than most procedures performed after that point.

Laboratory procedures that attempt to predict future conditions (eg. ABA) should ideally simulate the

Table 6. Multiple factor analysis of variance (ANOVA) for four laboratories, two samples (actually the same sample), and two replications.

	Main Effects			Interactions		
Parameter	Sample (S)	Lab (L)	Replicate (R)	S X L	S X R	L X R
pH	** †	NS	NS	**	NS	NS
Sand, %	NS	*	NS	NS	NS	NS
Clay, % (12-hr reading)	NS	*	NS	NS	NS	NS
Elec. Cond., mmhos/cm	NS	**	NS	NS	NS	NS
C.E.C., meq/100g	NS	**	NS	NDF	NDF	NDF
Neut. Potential, tkt	*	NS	NS	*	NS	NS
Total Sulfur, %	NS	**	NS	NS	NS	NS
Pyritic Sulfur, %	--‡	--	--	--	--	--
Exch. Acidity, meq/100g	NS	NS	NS	*	NS	NS
Acid/Base Account, tkt	**	NS	NS	EC	NS	EC
Total Cd, ppm	NS	*	NS	NS	NS	NS
Total Se, ppm	***	***	NS	**	NS	NS

† *, **, *** significant at the 0.05, 0.01, and 0.001 probability levels, respectively; NS - not significant at the p = 0.05 level; NDF - not enough degrees of freedom for interactions; and EC - interaction between main effects was confounded.

‡ - All pyritic-S values at or below the MDL (minimum detection level)

natural weathering that different materials will undergo in the post-mine minesoil, so that accurate predictions may be made of the probable soil conditions for reclamation. Unfortunately, it is not known how these materials will react in the field. Some of the shales will slake with a few wetting-drying cycles and disintegrate rapidly to the silt and clay fraction; other more cemented units (sometimes from the same geological unit) prove resistant to weathering and may endure as genuine gravel in the soils and should therefore not be milled to a fine powder in the laboratory.

Stewart (1996) found that current commercial laboratory methodology may not disperse clay minerals (in overburden materials) sufficiently, thereby resulting in lower clay content values. It is imperative, therefore, to ensure that flocculating and secondary cementing agents are removed and the sample is properly dispersed. This is accomplished by vigorously shaking the sample in a sodium hexametaphosphate solution overnight. Any carbonates, soluble salts, iron oxides, or organic constituents should also be removed from the sample. The intensity and duration of agitation can influence the results of particle size determinations.

Sample Extraction. There are several factors that may influence the amounts of constituents that are extracted; therefore, they should be controlled. These factors include, but are not limited to, the particle size of the sample, the amount of time the extractant is in contact with the sample, and the degree of sample-extractant agitation. It is possible to alter the results of many analyses by just altering one of the preceding variables.

Determinate and Indeterminate Errors. There are essentially four types of determinate errors that may affect data: instrument; method; personal bias; and gross mistakes (Skoog and West, 1980). These errors tend to be constant and proportional. Personal bias and instrument errors can be eliminated by training and proper calibration of instruments. Method errors can be identified through the analysis of standard samples. Gross errors can take the form of transcription errors, and may occur in the preparation of the final report, even if the laboratory procedures have been subject to a rigorous QA/QC program. Instances have occurred in which organic sulfur and pyritic sulfur values have been switched. These are very difficult to detect unless they can be correlated with other known geochemical information (e.g., whether from reduced overburden or from oxidized overburden).

Indeterminate errors can be caused by the cumulative effect of usually inconsequential differences (either from instrument, method, bias, or gross errors). One characteristic of indeterminate errors is that they occur randomly, thereby producing high results in some cases and lower results in others (Skoog and West, 1980).

Quality Assurance and Quality Control Considerations

The following section will only briefly cover the topic of QA/QC, since this issue is covered quite extensively in other literature. Quality assurance in a minesoil characterization program can include the use of standard reference materials (SRMs), split samples, and data quality assessment. These are procedures that provide assurance that the produced data meet defined standards of quality, with an associated level of confidence (Taylor, 1987). The goal of quality assurance is to identify, measure, and control the errors that may be introduced at any phase from sample collection to reporting of analytical results. Quality control is defined as an overall system of activities that provides data quality that is "satisfactory, adequate, dependable, and economic" (Taylor, 1987). The goal of quality control (quality improvement) is to minimize or correct for individual errors and their cumulative effect.

There are several terms that are used to assess measurement quality: bias, precision, and accuracy. Bias is a measurement of systematic error (deviation) in data. Precision (repeatability) is a measure of random variation in data; it indicates the reproducibility of a method (and can be expressed by the standard deviation). Accuracy describes the agreement between the amount of a component measured by a test method and the amount actually present.

A quality control program is essential in every laboratory, since quality control records provide proof of performance and offer data that can be referenced in the future. Laboratories, at a minimum, should keep an operating manual, a sample log, and a written record of all quality control checks (Bordner et al., 1978). Internal quality control can be achieved with control charts and multiple sample control charts. External quality control (among laboratories) is possible through the analysis of standard reference samples (APHA, 1985). The use of field duplicates, split samples, and standard reference materials during the course of a minesoil evaluation program is highly recommended.

It is uncertain how many state regulatory programs have implemented quality control guidelines with respect to overburden and minesoil data; however, Pennsylvania has included quality control procedures in its *Overburden Sampling and Testing Manual* (Noll et al., 1988), as part of its Small Operator Assistance Program. The quality control procedures in the manual address precision and accuracy; although parameters that have certified standard reference materials are the only ones checked for accuracy. Precision of analytical procedures is measured by repeatability intervals [or I(r)s]. Pennsylvania recommends that a minimum of one in twenty samples be a duplicate; a one in ten sample frequency is considered optimum. The target level of precision is dependent on the analytical parameter and the method used. The manual describes five methods of determining total sulfur and indicates the repeatability interval associated with each method (Table 7). The methodology used also affects the level of precision possible during fractionation of sulfur into the sulfate, pyrite, and residual (organic) fractions. Quality control procedures for pH and neutralization potential are also included in the manual. It would be recommended to investigate the rationale behind Pennsylvania's recommendations before a similar level of precision is initiated in another state, as there may be site-specific reasons for the criteria that have been implemented.

Conclusions and Recommendations

The results from the TMRA round robin program indicate that average inter-laboratory variability was approximately five times greater than intra-laboratory variability. The level of variation is affected by the sample being analyzed and the analytical parameter that is involved. The ranges in values from neutralization potential and total sulfur analyses had the lowest inter-laboratory to intra-laboratory (IIL) ratios, indicating the variability of the results may be due to inherent procedure variability or the heterogenous nature of the materials being analyzed. The larger inter-laboratory variability suggest that a greater convergence of analytical results may be attained through a rigorous examination of individual laboratory methodologies and implementa-

Table 7. Levels of accepted precision (repeatability) for duplicate samples, Pennsylvania's *Overburden Sampling and Testing Manual* (adapted from Noll et al., 1988).

Total Sulfur

Total Sulfur Method	Sulfur Content	Repeatability Interval [I(r)]
Eschka	% Sulfur < 2	± 0.05% S
	% Sulfur > 2	± 0.10 % S
Bomb Washing	% Sulfur < 2	± 0.05% S
	% Sulfur > 2	± 0.10% S
High Temperature Combustion, Infrared Absorption		I(r) = 0.03 ± 0.04x
Iodimetric Titration		I(r) = 0.08x
Acid Base Titration		I(r) = 0.06 + 0.03x

Sulfur Fractions

	Accepted Levels of Precision			
	EPA Method (Sobek et al., 1978)		ASTM Method (D2492)	
Sulfur Fractions	% Sulfur < 2	% Sulfur > 2	% Sulfur < 2	% Sulfur > 2
Sulfate	± 0.05% S	± 0.10% S	± 0.02% S	± 0.02% S
Pyritic	± 0.15 % S	± 0.25% S	± 0.05% S	± 0.10% S
Organic	± 0.05% S	± 0.05% S	± 0.10% S	± 0.20% S

pH	Accepted Precision, std. units
	± 0.3

Neutralization Potential	Range of NP Values, tkt	Accepted Precision for
	-30	± 5
	-70	± 10
	-400	± 25
	> 500	± 50

tion of widely-used standard operating procedures (SOPs).

An examination of coefficient of variation (C.V.) data from twelve analytical parameters indicated that pH is the least variable and trace elements are among the most variable. The inclusion of near-minimum detection level (MDL) values tended to increase the C.V. of data.

The ANOVAs indicated that the sources of variation of most parameters were inconsistent over time. The differences between intra-laboratory data (replicates) were not a significant source of variation. Differences in equipment, sample extraction methods, and personnel could account for different values amont the laboratories. It would be necessary, in cases where there were differences among laboratories, to ensure that all of the laboratories were, in fact, following the same procedures. Analytical parameters that exhibited a sample by laboratory (S X L) interaction suggest that there was inconsistency between laboratories over time. Results of samples M and P indicate that there was either a significant laboratory main-effect or S X L interaction for most of the analytical parameters, implying that either lab inconsistencies or sample heterogeneity could be the cause. The use of standard reference materials (SRMs), with known parameter values, can help identify which laboratories are accurate.

There are analytical considerations that require attention, such as sample preparation and pretreatment, sample extraction, and both determinate and indeterminate errors. The authors' previous experience indicates that sample preparation may be a critical step where variability can be introduced. Any changes in methodology, that involve the previous factors, can affect analytical results. The development and implementation of standard operating procedures is an essential process for any laboratory. The variable results obtained by the TMRA round robin program illustrate the need for QA/QC procedures and the use of good laboratory practices. The use of field duplicates, split samples, and standard reference materials is recommended for any sampling program.

The program's results also prompted the following questions: 1) are the levels of variability of sufficient importance to warrant additional research and 2) are there practical mine reclamation implications to the variability exhibited in the round robin data?

References

APHA (American Public Health Association). 1985. Standard methods for the examination of water and wastewater. 16th ed. Washington, D.C.

ASTM (American Society for Testing and Materials). 1984. Standard method of testing for forms of sulfur in coal. Annual Book of Standards. D2492. 84:323-327. Philadelphia, PA.

Bajo, Sixto. 1978. Volatilization of arsenic (III, V), antimony (III, V), and selenium (IV, VI) from mixtures of hydrogen fluoride and perchloric acid solution: Applications to silicate analysis. Analytical Chemistry, vol. 50:649.
https://doi.org/10.1021/ac50026a028

Barth, D.S. and B.J. Mason. 1984. Soil sampling quality assurance user's guide. EPA-600/8-84-043. Environmental Research Center, University of Nevada, Las Vegas. Las Vegas, NV.

Black, C.A. (Editor-in-Chief). 1965. Methods of soil analysis. Chemical and microbiological properties, Part 2. Agronomy Monograph No. 9. American Society of Agronomy, Inc. Madison, WI.

Bordner, R., J. Winter, and P. Scarpino (eds.). 1978. Microbiological methods for monitoring the environment: Water and wastes. U.S. Environmental Protection Agency. EPA-600/8-78-017. Cincinnati, OH.

EPA (Environmental Protection Agency). 1995. USEPA contract laboratory program: Statement of work for inorganics analysis, multi-media concentration. ILMO 4.0. Publication 9240.1-30; EPA/540/R95/121. Office of Solid Waster and Emergency Response. U.S. Environmental Protection Agency. Cincinnati, OH.

EPA (Environmental Protection Agency). 1979. Methods for the chemical analysis of water and wastes. EPA 600/4-79-020. Industrial Environmental Research Laboratory, U.S. Environmental Protection Agency. Cincinnati, OH.

Manugistics, Inc. 1992. STATGRAPHICS® Plus, Version 6.1. Rockville, MD.

Munk, L.P., F. Rivera, and S.E. Fisher, Jr. 1996. Developing efficient and defensible soil and overburden characterization plans. *In* Proc.: Planning, Rehabilitation and Treatment of Disturbed Lands, Billings Symposium. March 17-23, 1996, Billings, MT.

Noll, D.A., T.W. Bergstresser, and J. Woodcock. 1988. Overburden sampling and testing manual. Prepared under contract by Earthtech and Geochemical Testing Divisions of Energy Center, Inc.; Contract No. ME 86120. Small Operator Assistance Program, Bureau of Mining and Reclamation, Dept. of Environmental Resources. Harrisburg, PA.

Railroad Commission of Texas. 1996. Advisory Notice AG-RP-145: Overburden and minesoil samples - Preparation procedure for large and small volume samples. Railroad Commission of Texas, Surface Mining and Reclamation Division. Austin, TX.

Railroad Commission of Texas. 1989. Overburden parameters and procedures. Railroad Commission of Texas, Surface Mining and Reclamation Division. Austin, TX.

Railroad Commission of Texas. 1988. Technical Release No. SA-2, Materials suitable forplacement in the top four feet of leveled minespoil, including topsoil substitutes. Railroad Commission of Texas, Surface Mining and Reclamation Division. Austin, TX.

Skoog, D.A. and D.M. West. 1980. Analytical chemistry. 3rd ed. Saunders College Publishing. Philadelphia, PA.

Sobek, A.A., W.A. Schuller, J.R. Freeman, and R.M. Smith. 1978. Field and laboratory methods applicable to overburdens and mine soils. EPA 600/2-78-054. U.S. Environmental Protection Agency. Cincinnati, OH.

Stewart, W.M. 1996. Factors affecting the ratio of cation exchange capacity to clay content in east Texas lignite overburden. Ph.D. Dissertation, Texas A&M University. College Station, TX.

Taylor, J.K. 1987. Quality assurance of chemical measurements. Lewis Publishers, Inc. Chelsea, MI.