



# Standard Guide to the Evaluation of Measurements Made by On-Line Coal Analyzers<sup>1</sup>

This standard is issued under the fixed designation D 6543; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon ( $\epsilon$ ) indicates an editorial change since the last revision or reapproval.

<sup>e1</sup> NOTE—Figs. 7–10 were editorially corrected in December 2000.

## 1. Scope

1.1 This guide provides techniques to be used for the evaluation of the measurement performance of on-line coal analyzers.

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

## 2. Referenced Documents

### 2.1 ASTM Standards:

- D 121 Terminology of Coal and Coke<sup>2</sup>
- D 2013 Method of Preparing Coal Samples for Analysis<sup>2</sup>
- D 2234 Test Methods for Collection of a Gross Sample of Coal<sup>2</sup>
- D 4702 Guide for Inspecting Crosscut, Sweep Arm, and Auger Mechanical Coal Sampling Systems for Conformance with Current ASTM Standards<sup>2</sup>
- D 6518 Practice for Bias Testing Mechanical Coal Sampling Systems
- E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods<sup>3</sup>
- E 178 Practice for Dealing With Outlying Observations<sup>3</sup>
- E 456 Terminology Relating to Quality and Statistics<sup>3</sup>

## 3. Terminology

### 3.1 Definitions:

3.1.1 *analyzer system, n*—a coal quality measurement system which includes an on-line coal analyzer and which may include one or more stages of a coal-sampling system.

3.1.2 *calibration, n*—mathematical modeling of analyzer and comparative coal sampling and analysis data. Factors from the model are used in the on-line analyzer control software.

3.1.3 *Grubbs estimator, n*—an estimate of the measurement precision of an on-line analyzer **(1-3)**.<sup>4</sup>

3.1.4 *on-line analyzer, n*—an analytical tool consisting of an instrument and systems, which together provide measurements, or estimates, or both, of coal quality parameters.

3.1.5 *outlier, n*—an extreme value that statistical tests indicate to be far enough from other results in a population under consideration to cause suspicion that the value is not a member of the population.

3.1.6 *reference material, n*—material of stable composition that may be used to generate static analyzer measurements.

3.1.7 *reference system, n*—a measurement system used to measure the characteristics of a lot of coal that are also measured by an on-line-analyzer, and against which the on-line analyzer measurements are compared.

3.1.8 *standardization, n*—calibration of an instrument to a reference material using static stability measurements.

3.1.9 *static stability, n*—an estimate of the measurement precision of an instrument obtained on material that is not moving. The estimate normally is expressed as the standard deviation and average of the measurements for a given period of time.

3.1.10 *synchronization error, n*—an error that occurs from comparing measurements made by an on-line analyzer and a reference system that are not measuring exactly the same lot because of temporal and/or spatial offsets.

## 4. Summary of Guide

4.1 This guide outlines the evaluation of the measurement performance of an on-line analyzer using comparative measurements. The comparative test uses a paired comparison of analysis from a reference method using ASTM sampling, sample preparation, and analysis methods for a lot of coal with the analysis from the on-line analyzer for the same lot of coal. The data resulting from the comparative test may be evaluated using graphical and statistical techniques outlined below.

<sup>1</sup> This guide is under the jurisdiction of ASTM Committee D05 on Coal and Coke and is the direct responsibility of Subcommittee D05.23 on Sampling.

Current edition approved April 10, 2000. Published May 2000.

<sup>2</sup> *Annual Book of ASTM Standards*, Vol 05.06.

<sup>3</sup> *Annual Book of ASTM Standards*, Vol 14.02.

<sup>4</sup> The boldface numbers in parentheses refer to the list of references at the end of this standard.

4.2 Various techniques are recommended by on-line analyzer manufacturers for standardization or static testing. These techniques are useful for establishing a benchmark before conducting a comparative test. These techniques also may be used as diagnostic tests in accordance with methods recommended by on-line analyzer manufacturers and graphical and statistical techniques included in this guide.

## 5. Significance and Use

5.1 On-line analyzers are used to provide quality data on lots of coal. The resulting quality data are used as a production tool or for some contractual application. This guide provides means of evaluating the system and data produced.

5.2 The user should become familiar with the document's terminology and layout. The section on test design and data collection will provide the means by which all the analysis data will be gathered. The test design should be carefully considered so as to assure meeting the user's requirements.

5.3 The procedures defined in this guide can be used to test the accuracy and precision of an on-line analyzer, for acceptance after its installation, to check precision and accuracy during routine use (quality control), when changes are made to the system, when the nature of the coal being tested changes, and to determine mathematical factors to be used for calibration of the on-line analyzer.

## 6. Selection and Conduct of Performance Evaluations

### 6.1 Introduction:

6.1.1 Several techniques can be used to evaluate the performance of an on-line analyzer. These techniques provide data that can be evaluated by using the graphical and statistical methods described in Section 7 of this guide.

6.1.2 At time of installation, all of the graphical and numerical methods outlined in this may prove useful. On a routine basis, conducting any of the instrument stability checks and comparative evaluations that do not disrupt normal operations may prove useful. Control charts may be applied to all the performance measures that are gathered on a routine basis, for example, mean analysis value of reference material, RMSD, and so forth.

6.1.3 In the event that there is a change to the operational parameters or the equipment associated with the analyzer or the reference system, comparative checks should be performed. In addition to comparative checks, standardization or static checks, or both, as recommended by the on-line analyzer manufacturer may be helpful.

6.1.4 Changes in the following may impact performance evaluations. Coal characteristics, such as particle size, source of coal, mining techniques, degree of preparation, and so forth, which if changed from previous test periods and are not in the analyzer calibration database, may affect analyzer precision and accuracy:

- 6.1.4.1 Analyzer components;
- 6.1.4.2 Coal-handling system;
- 6.1.4.3 Laboratory services;
- 6.1.4.4 Sampling technique;
- 6.1.4.5 Coal flow rates; and
- 6.1.4.6 Power disturbances.

6.2 *Static Stability Measurements for Baseline Assessment*—A reference material may be used to provide a baseline assessment of static measurement precision. The reference material may be used to compare current mean and standard deviation values with previous mean and standard deviation values, previously collected in the same manner. The resulting comparative data may be used to assess whether any change in the mean or standard deviation of the static measurements may be attributed to a change in the operating characteristics of the analyzer, in the absence of the influence of sampling and analysis.

6.2.1 Reference materials may include actual coal in a sealed container, cast high-carbon solids, or other materials which may be provided by the analyzer manufacturer. The primary characteristic of these materials is that they not be subject to changes in composition.

6.2.2 The results of this evaluation can indicate whether analyzer performance has significantly degraded or whether a shift may have occurred. If so, it may be possible to adjust the analyzer to restore initial performance. If the user wishes only to measure current static repeatability, any available coal may be used in the analysis zone of the analyzer. Note, however, that the actual standard deviation in static repeatability tests might be influenced by the composition of the coal being examined or analyzer factors, such as the strength of the radioactive sources used by the analyzer or condition of analyzer electronic components.

6.2.3 It is essential that the length of the analysis period be defined, for example, 1, 2, or 5 min, and be constant in the static stability test. The static standard deviation resulting from the static stability test decreases as the length of the analysis period increases. Static stability testing may be conducted on the same day(s) as comparative tests. One type of static test involves a comparison of the analyzer to static coal over many hours.

### 6.3 Comparison of Analyzer System to Reference System Measurements:

6.3.1 Once an analyzer installation has been completed and calibration adjustments have been made, the analyzer owner may require acceptance testing. Also, the analyzer owner may decide to relocate the analyzer. In these cases, comparison tests, that is, to compare the analyzer system results to conventional sampling and analysis techniques, will provide the user calibration verification data and/or data that could be used for recalibration of the analyzer.

6.3.2 Depending upon the comparative technique chosen, there may be need for more than one comparison. These comparisons may be conducted in a batch over several hours or days or continuously throughout the operational life of the analyzer system.

6.3.3 If two independent conventional coal-sampling and laboratory analysis measurements can be made from each of a series of batches of coal interrogated by the analyzer, methods developed by Grubbs (**1-3**) can be used to provide unbiased estimates of the measurement precision of the analyzer and of the conventional sampling and analysis methods.

6.3.4 Any two series of measurements are independent if their measurement errors are uncorrelated. Correlation of

measurement errors can be avoided and independence assured by use of a true random selection of physical increments or samples of material or by using different schemes and equipment for collection, preparation, and laboratory analysis of the samples, or both. A complete treatment of the subject of independence of measurements and the various means of assuring independence is beyond the scope of this guide.

6.4 Sampling Considerations:

6.4.1 Selection of Appropriate Sampling and Sample Preparation Methods—The following increment collection methods (see Test Methods D 2234) are listed in general order of preference; this preference is not rigid. Often, practical considerations may supercede increment selection decisions. The method to be used for sample preparation should be determined before the beginning of increment collection. Sample preparation techniques should remain consistent (see Method D 2013). Before installation of the analyzer, consideration should be given to the ability to obtain representative samples for comparison to analyzer measurements and the regimen for sample handling and analysis. For the use of mechanical sampling systems, inspection (see Guide D 4702) and bias testing (see Practice D 6518) are good methods for evaluation of the system’s suitability for the test.

6.4.1.1 Full Stream Cuts—Whenever possible, the full stream cut method of increment collection should be used in the evaluation of an on-line analyzer. This is imperative for drawing inference using statistical techniques.

6.4.1.2 Stopped Belt Cut—Although using “stopped belt” increments allows drawing statistical inferences from the data, the stopped belt increment has limited applicability in the evaluation of on-line analyzers. This especially is true in collection of a sample from the lot interrogated by an analyzer operating in conjunction with a mechanical sampling system. The stopped belt increment may have applicability in instances in which the analyzer is used in conjunction with mechanical sampling, and the discharge of the analyzer may be metered onto a belt that is subject to stopping without interference with the analyzer’s operation. The stopped belt increment may be useful in evaluating analyzers that are not associated with mechanical sampling and that are being evaluated for very short-term performance, that is, less than 1-min results. In such instances, each stop of the belt is used to collect a sample from a relatively long portion of the belt. As an example, the analyzer might collect data for 30 s and the belt stopped in such a manner as to allow access to the entire segment of material interrogated by the analyzer (typically, 300 ft for a belt running 600 ft/min for 30 s). A sample may now be collected by removing increments from the stopped belt throughout the entire stopped portion. This method reduces the number of belt stops required to assemble a database but is limited in applicability by the short analysis time and physical requirements of the facility wherein the equipment is installed. Nevertheless, this technique may be considered when no reliable mechanical technique is available and the alternative is a manual part stream cut.

6.4.1.3 Part Stream Increments—This mode of sample collection may be required in the evaluation of analyzers not associated with mechanical sampling or in some tests in which

more than one comparison will be made. Since this type of sample does not allow statistical inferences to be drawn in a rigorous manner, its use is discouraged. Indeed, the only use of this type of comparison might be for control charting purposes relative to an historical reference.

6.4.2 Selection of Reference Sample Point(s):

6.4.2.1 Comparative tests can be used to evaluate the performance of either the analyzer itself or the “analyzer system” (consisting of the analyzer and the sampling system that feeds it). The comparative evaluations of the analyzer system can be used to determine the ability of the analyzer system to measure the characteristics of the main coal stream. The selection of the reference system sample point(s) determines whether the comparative tests will assess the analyzer or the analyzer system.

6.4.2.2 The most direct and practical two-instrument test, when the analyzer is fed the secondary reject of a mechanical sampling system, uses the final save to compare directly to the analyzer. A manual or mechanical sample collected from the analyzer discharge may provide an independent sample, which may be used to assess the performance of the analyzer (see Figs. 1-3).

6.4.2.3 There are instances when there is no save sample (see Fig. 4) associated with the system feeding the analyzer. In this case, it is possible to construct a test with several comparisons by collecting separate samples from the analyzer feed and discharge. Since independence of the samples is important for three instrument evaluations, subsequent preparation and analyses should be performed at different times by different parties.

6.4.2.4 In some instances, the discharge of the analyzer may be fed to further stages of mechanical sampling. A single stage of sampling downstream of the analyzer is most common (see Fig. 5). In this case, the secondary save will provide a convenient comparison sample. A test of such a system that requires more than one set of comparative data might incorporate a series of samples (stopped belt or full stream cut) collected from the analyzer discharge or the secondary reject stream, as well. In the relatively rare circumstances in which the analyzer discharge feeds two additional stages of sampling, the tertiary save is recommended for two-instrument and bias testing, and a stopped belt or full stream cut of the secondary or tertiary rejects of the analyzer discharge may be used as a third instrument. Practical considerations of increment collection at the secondary reject should be balanced with considerations of sampling variances introduced by crushing and

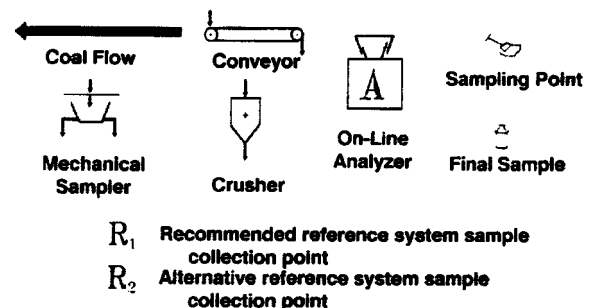


FIG. 1 Key to Schematics

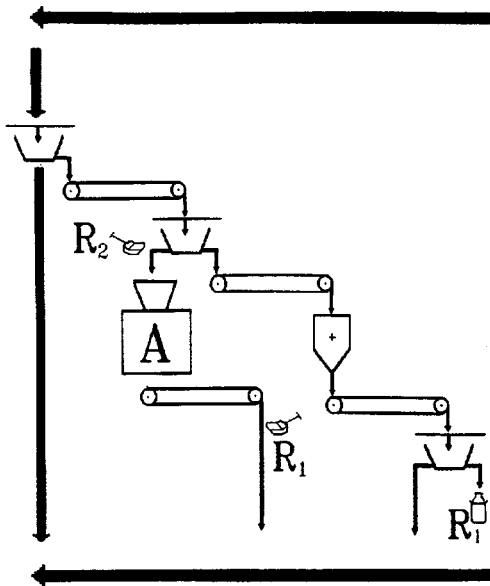


FIG. 2 Analyzer on Secondary Reject—Three-Stage Mechanical Sampling

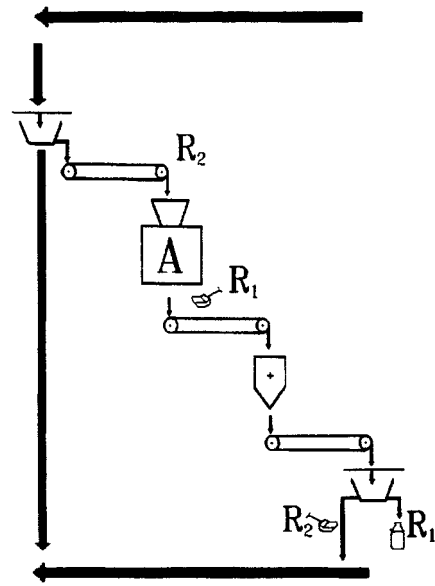


FIG. 4 Analyzer on Primary Save—Two-Stage Mechanical Sampling

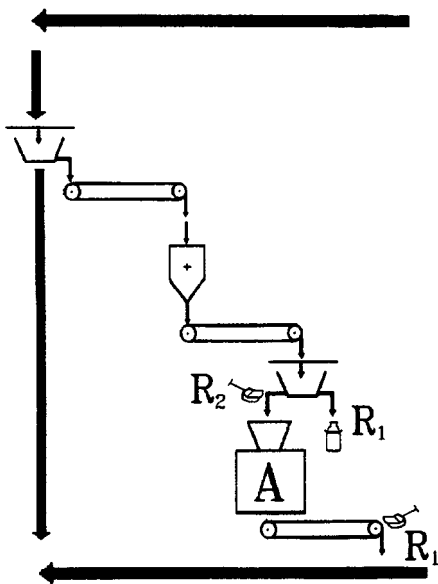


FIG. 3 Analyzer on Secondary Reject—Two-Stage Mechanical Sampling

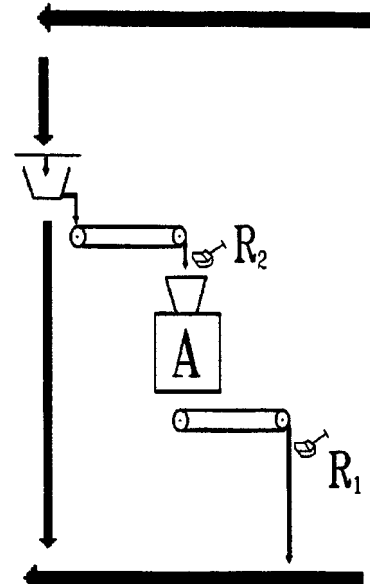


FIG. 5 Single-Stage Mechanical Sampling

tertiary sampling. For two instruments to be independent of each other, one or both instruments must interrogate the stream of interest without changing the characteristics of the stream. This may be true in some through-belt noncontacting configurations. In the case of flow-through analyzers that require a sample, the independence of systems is obtained in a case in which the primary coal stream is sampled by one instrument before being sampled by another instrument (see Fig. 6). In this case, the two systems may be evaluated by comparing the analyzer values to the final save of the mechanical system. The analyzer itself may be evaluated by comparison to samples collected at its discharge.

6.4.3 *Synchronization*—To evaluate the analyzer's performance, the lot of coal interrogated by the analyzer must be the

same lot of coal measured by the reference system. Generally, this is accomplished by sampling the coal stream after it passes the analyzer's analysis zone. Care should be taken to ensure that transport times within the coal-handling system of interest, within mechanical sampling systems used to provide comparative measurements, and between sample collection and analyzer be measured and accounted for in any comparisons. Comparisons in which intervening storage of the coal takes place between the analyzer and reference measurements should be avoided.

6.4.4 *Length of Comparison Period*—The length of time chosen for a comparison period should be commensurate with the period of time, that is, the lot time, during which the analyzer will normally produce a result of interest to the user.



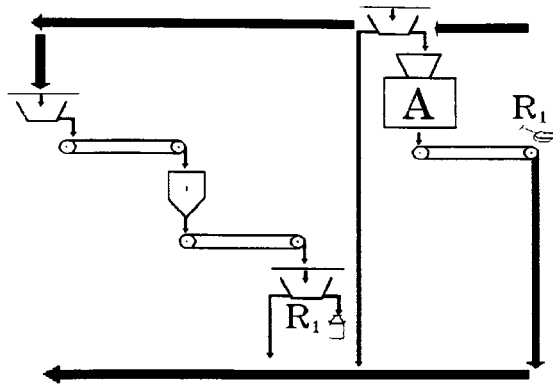


FIG. 6 Analyzer Distinct from Reference

6.4.5 *Number of Comparisons*—Although as few as 15 comparisons may be valuable in establishing statistical control for conventional control charting, as many as 60 comparisons may be justified if one is to obtain reasonably close confidence limits for the Grubbs estimate of precision.

## 7. Evaluation of Data

7.1 *Introduction*—Once comparative data have been collected, it is recommended that the data be evaluated using both graphical and statistical, that is, numerical techniques. The sections that follow give recommendations for the construction and interpretation of various charts or graphs as well as the techniques used to compute and interpret various statistics.

### 7.2 Stability Evaluations:

#### 7.2.1 Graphical Techniques, see 7.3.2.

7.2.2 *Numerical Methods and Their Interpretation*—The two statistics most relevant to static stability evaluations are the average of the static analyses and the standard deviation of the static analyses. To calculate the average of the static analyses and the standard deviation of the static analyses, the individual analyzer static analysis data points (analysis by analysis) are needed. The sample average is given by the formula:

$$\bar{x} = [\sum_{i=1}^n x_i]/n \quad (1)$$

where:

$\bar{x}$  = the average,

$x_i$  = the *i*th measurement, and

*n* = the number of measurements.

The sample standard deviation  $s_d$  is given by:

$$s_d = \sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2]/(n-1)} \quad (2)$$

7.3 *Comparative Evaluations*—Comparison of analyzer results to those generated by conventional techniques should be performed and assessed using the techniques described below. These evaluations require a series of comparisons between analyzer measurements and reference system measurements. It is recommended that such comparisons be made routinely.

### 7.3.1 Outliers:

7.3.1.1 At times, a measurement from an analyzer or reference system may appear to be an aberrant value. Statistical rules leading one to look for causes of outliers when they exist are given in Practice E 178. In applying the methods of

Practice E 178, the difference between the analyzer and reference value is computed for each pair.

7.3.1.2 Each difference is then divided by the sample standard deviation of the differences, with the absolute value of the quotient designated as *U<sub>i</sub>* for the *i*th pair. It is recommended here that any pair (*i*), whose *U* value is found using Table 1 of Practice E 178 to exceed the table value for an appropriate upper significance level, that is, 99 %, be treated as an outlier. An investigation of the cause of the outlier should be undertaken, and it may be appropriate to exclude the outlier data from analyzer performance evaluation calculations.

7.3.2 *Graphical Methods*—Regardless of the numerical method(s) used to assess instrument performance, any comparison between analyzer values and reference system values should include graphical displays. At a minimum, a sequential plot and an *x-y* plot should be constructed. Many conclusions can be drawn or hypotheses constructed about the quality of the calibration of the instrument, the synchronization of the data, and the precision of the analyzer or the reference system.

7.3.2.1 *Sequential Plot*—Fig. 7 is an example of a sequential plot, showing the analyzer and reference system measurements versus the lot number on the same graph. Provided that product variation is sufficiently large compared to the precision of each of these measurements, one expects to see “tracking.” That is, when the conventional measurement shows a decrease in ash, one expects the analyzer measurement to show a similar decrease. Fig. 7 below appears to be an example of good tracking. A sequential plot of the difference between the analyzer and the reference sample result also may be used. Shown in Fig. 8 is a magnification of the Fig. 7 plot for lot numbers 10 to 20. Notice that, on this scale, the “tracking” is less convincing. Because the product variability is quite small during this time, this does not necessarily mean poor performance. Again, tracking is expected only when product variability is large compared to the measurement precision of each system.

7.3.2.2 *x-y Plot*—Shown in Fig. 9 is an example of an *x-y* plot or “scatter-plot” of analyzer results versus reference system measurements (the data are the same as those from Fig. 7). For an accurate and well-calibrated analyzer, one expects the points to fall along a line with zero intercept and a slope of one, as indicated in the graph. An improperly calibrated analyzer may show a good linear pattern but with a nonzero intercept or a non-unity slope in the parameter being measured, while an imprecise analyzer will show a large scatter. Again, unless the product variability is reasonably large compared to instrument precision, comparisons may not be very conclusive.

NOTE 1—Fig. 9 shows a reasonable scatter around the 45° line (a line with zero intercept and a slope of one), with the exception of three potential outliers at the upper right. The numerical methods discussed in this guide would not by themselves indicate a problem of this sort.

### 7.3.3 Numerical Methods:

#### RMSD Accuracy Statistic

7.3.3.1 The root mean squared difference (RMSD), is as its name implies the square root of the mean of the differences between reference and analyzer values squared.

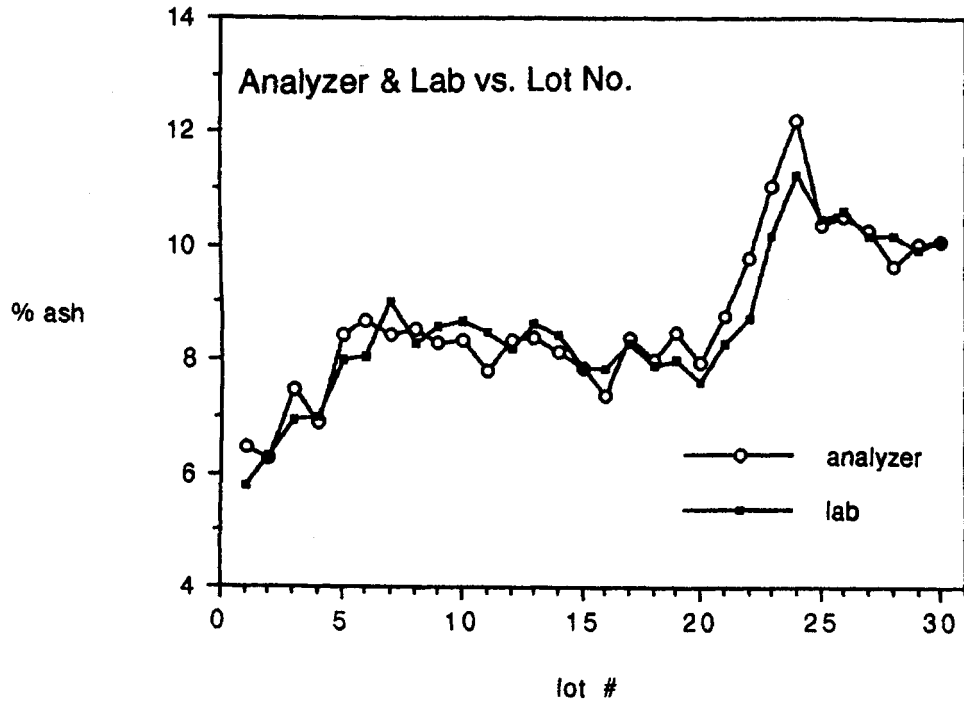


FIG. 7 Sequential Plot Showing the Analyzer and Reference System Measurements Versus the Lot Number on the Same Graph

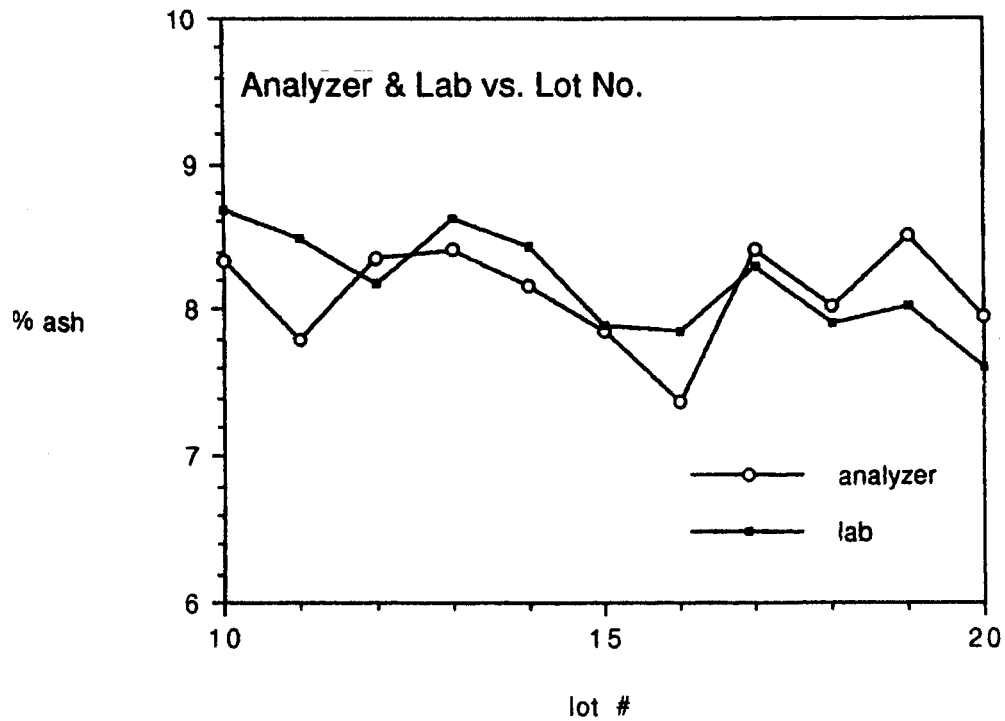


FIG. 8 Magnification of Fig. 7 Plot for Lot Numbers 10 to 20

7.3.3.2 Calculations—By formula, RMSD is calculated as follows:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (3)$$

where:

- $n$  = the number of paired comparisons,
- $x_i$  = the analyzer value in the  $i$ th comparison, and
- $y_i$  = the reference value in the  $i$ th comparison.

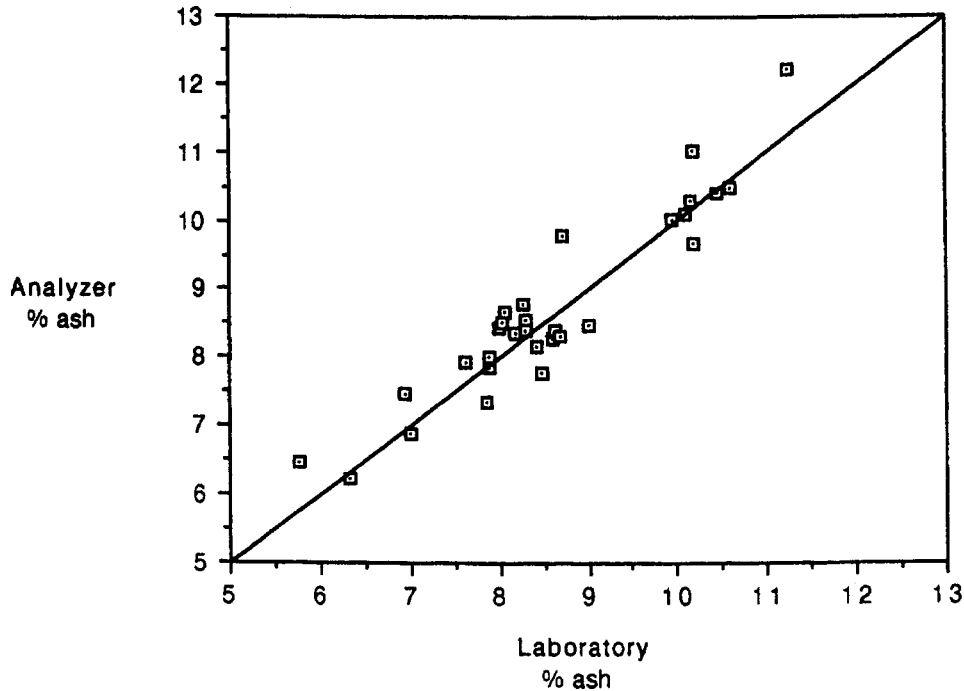


FIG. 9 x-y Plot or “Scatter-Plot” of Analyzer Results Versus Reference System Measurements (the Data are the Same as Those from Fig. 7)

7.3.3.3 *Interpretation*—Note that the RMSD statistic becomes larger (or worse) as any of the following occur: the bias increases, the precision of the instrument worsens, or the precision of the reference system worsens. A good (or small) RMSD statistic, therefore, usually implies a precise analyzer with small bias being evaluated by a fairly precise reference system. On the other hand, a small RMSD may arise strictly because the coal quality values have stayed within a narrow range during the evaluation period. A large RMSD does not necessarily mean the analyzer is performing poorly. The large RMSD may result from a poor calibration that leaves the instrument biased. See Fig. 10 as an example of a very precise instrument whose RMSD is large because of improper calibration. The proper response to this assessment of analyzer performance should be to adjust the calibration, not to reject the instrument as unworthy. Furthermore, a perfect instrument may produce a sizable RMSD if the ability of the reference system to equal the true value of the coal quality measurement is poor. This situation could arise if the variances of sampling, preparation, or analysis in the reference system are significant.

**Grubbs Estimators of Measurement Precision**

7.3.3.4 After attaining the analyzer’s measurement and two independent reference measurements for each of a recommended number of 60 or more batches of coal, the Grubbs estimators of measurement precision can be calculated (see Figs. 1-9).

7.3.3.5 *Calculations:*

Define:

- $s_x^2$  = sample estimate of the variance of analyzer observations,
- $s_y^2$  = sample estimate of the variance of Ref (1) observations,
- $s_z^2$  = sample estimate of the variance of Ref (2) observations,
- $s_{x-y}^2$  = sample estimate of the variance of differences between analyzer and Ref (1) observations,
- $s_{x-z}^2$  = sample estimate of the variance of differences between analyzer and Ref (2) observations,
- $s_{y-z}^2$  = sample estimate of the variance of the differences between Refs (1 and 2) and observations, and
- $s_{x+y+z}^2$  = sample estimate of the variance of the sums of analyzer, Refs (1 and 2) observations:

where:

$$s_d = \left[ \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 \right]^{0.5} \tag{4}$$

and where  $s_d^2$  is the sample variance of differences,  $d_i = x_i - y_i$  and

$$\bar{d} = \left[ \sum_{i=1}^n d_i \right] / n \tag{5}$$

Grubbs (1) shows the following variance estimates to be unbiased maximum likelihood estimates:

$$V_c = \left[ s_{x+y+z}^2 - \frac{1}{2}(s_{x-y}^2 + s_{x-z}^2 + s_{y-z}^2) \right] / 9 \tag{6}$$

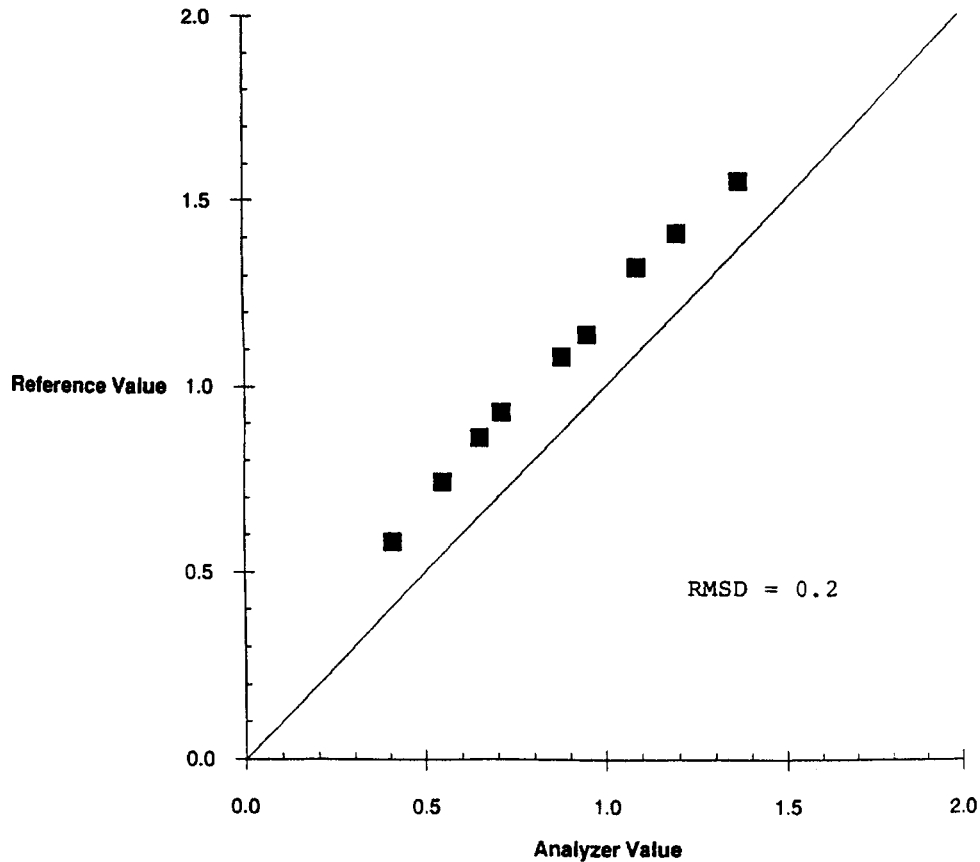


FIG. 10 Example of a Very Precise Instrument Whose RMSD is Large Because of Improper Calibration

$$V_a = (s_{x-y}^2 + s_{x-z}^2 - s_{y-z}^2)/2$$

$$V_1 = (s_{x-y}^2 + s_{y-z}^2 - s_{x-z}^2)/2$$

$$V_2 = (s_{x-z}^2 + s_{y-z}^2 - s_{x-y}^2)/2$$

where:

- $V_c$  = the variance of the coal, batch by batch;
- $V_a$  = the Grubbs estimate of the variance of the errors of batch measurement by the analyzer;
- $V_1$  = the estimate of the variance of the errors of batch measurement by the Ref (1) method; and
- $V_2$  = the estimate of the variance of the errors of batch measurement by the Ref (2) method.

7.3.3.6 The estimate of the index of measurement precision for each of the three systems is given by:

$$\begin{aligned} s_a &= \sqrt{V_a} \\ s_1 &= \sqrt{V_1} \\ s_2 &= \sqrt{V_2} \end{aligned} \tag{7}$$

7.3.3.7 One may wish to determine whether there is sufficient evidence to conclude that the Grubbs calculated indices of precision are different from values of interest at a particular level of confidence. If this is the case, the user is referred to Refs (3-5).

7.3.3.8 Interpretation—The Grubbs Estimator Method is used to estimate the precision of three independent measurements of the same property without bias. One criterion for this test is that the three instruments used to gather the data are

independent. A basic knowledge that the three reference systems are not grossly different in precision is advisable. This knowledge may be useful during the design of the test logistics. If one system is grossly different in precision than the other two, then a negative variance may result. The result of a negative variance does not mean that the test result is not useful (5). The use of the Grubbs estimator in this case, has been discussed (6). The negative Grubbs estimator could mean that one reference method, as a result of an inadequate test design or conduct of the test, may have a large variance in comparison to the other reference method. This could also mean that the precision of the analyzer was very low compared to the precision of the two reference methods.

7.3.3.9 Standard Deviation of the Differences—The sample standard deviation of the differences between the analyzer and reference measurements is an indicator of the combined indexes of precision of both the measurement systems.

7.3.3.10 Calculations—The sample standard deviation of the differences  $SD_d$  is given below. Note that with the reasonable assumption that the analyzer measurement errors are uncorrelated with reference measurement errors.

where  $d_i = x_i - y_i$  and

$$\bar{d} = [\sum_{i=1}^n d_i]/n \tag{8}$$

7.3.3.11 Interpretation—If the variance of reference measurement errors is small compared to the variance of analyzer measurement errors, and if the number of data sets is 30 or



more, the sample standard deviation of the differences may reasonably approximate an upper bound for the analyzer precision.

7.3.3.12 *Test for Bias*—Bias of interest in the application of on-line analyzers is the bias between analyzer measurements and reference measurements made by conventional coal sampling and analysis. The statistical techniques in the Practice

D 6518 may be used to test for bias between the measurements made by an on-line analyzer and reference measurements made by conventional coal sampling and analysis.

**8. Keywords**

8.1 comparative measurement; on-line coal analyzer

**ANNEXES**

**(Mandatory Information)**

**A1. EXAMPLE CALCULATIONS, THREE-INSTRUMENT GRUBBS TEST**

A1.1 Table A1.1 illustrates hypothetical data from a three-

**TABLE A1.1 Example of Three-Measurement Test for Precision Estimates**

| Batch No.                  | Analyzer (x)                          | Ref (1) (y)                           | Ref (2) (z)                           | x-y                           | x-z                           | y-z                           | x+y+z                            |
|----------------------------|---------------------------------------|---------------------------------------|---------------------------------------|-------------------------------|-------------------------------|-------------------------------|----------------------------------|
| 1                          | 12.75                                 | 12.01                                 | 12.27                                 | 0.74                          | 0.48                          | -0.26                         | 37.03                            |
| 2                          | 11.35                                 | 10.99                                 | 11.24                                 | 0.36                          | 0.11                          | -0.25                         | 33.58                            |
| 3                          | 15.92                                 | 15.35                                 | 15.83                                 | 0.57                          | 0.09                          | -0.48                         | 47.10                            |
| 4                          | 10.48                                 | 10.83                                 | 10.97                                 | -0.35                         | -0.49                         | -0.14                         | 32.28                            |
| 5                          | 12.32                                 | 12.45                                 | 11.95                                 | -0.13                         | -0.37                         | 0.50                          | 36.72                            |
| 6                          | 13.14                                 | 13.05                                 | 12.60                                 | 0.09                          | 0.54                          | 0.45                          | 38.79                            |
| 7                          | 15.26                                 | 15.43                                 | 16.10                                 | -0.17                         | -0.84                         | -0.67                         | 46.79                            |
| 8                          | 13.40                                 | 14.30                                 | 13.75                                 | -0.90                         | -0.35                         | 0.55                          | 41.45                            |
| Variance (S <sup>2</sup> ) | 3.3376<br>S <sub>x</sub> <sup>2</sup> | 3.3044<br>S <sub>y</sub> <sup>2</sup> | 3.8760<br>S <sub>z</sub> <sup>2</sup> | 0.2833<br>s(x-y) <sup>2</sup> | 0.2495<br>s(x-z) <sup>2</sup> | 0.2247<br>s(y-z) <sup>2</sup> | 30.7966<br>s(x+y+z) <sup>2</sup> |

measurement test. Keep in mind that although only  $n = 8$  sets of data are used in the illustration, useful tests may require a number of data sets with  $n$  equal to or greater than 60.

A1.2 For the data of Table A1.1 (see 7.3.3.5):

$$V_c = [30.7966 - (1/2)(0.2833 + 0.2495 + 0.2247)]/9 = 3.3798$$

$$V_a = (0.2833 + 0.2495 - 0.2247)/2 = 0.1540$$

$$V_1 = (0.2833 + 0.2247 - 0.2495)/2 = 0.1292$$

$$V_2 = (0.2495 + 0.2247 - 0.2833)/2 = 0.0954$$

and the estimates of precision are:

$$s_a = (0.1540)^{1/2} = 0.39$$

$$s_1 = (0.1292)^{1/2} = 0.36 \text{ and}$$

$$s_2 = (0.0954)^{1/2} = 0.31$$

**A2. EXAMPLE CALCULATIONS: RMSD AND S<sub>d</sub> CALCULATIONS**

A2.1 An example is shown in Table A2.1.

**TABLE A2.1 Ash Measurement (Wt., %)**

| Comparison Number | Analyzer | Reference <sup>A</sup> | Difference | (Difference) <sup>B</sup> |
|-------------------|----------|------------------------|------------|---------------------------|
| 1                 | 12.75    | 12.14                  | 0.61       | 0.372                     |
| 2                 | 11.35    | 11.12                  | 0.23       | 0.053                     |
| 3                 | 15.92    | 15.59                  | 0.33       | 0.109                     |
| 4                 | 10.48    | 10.90                  | (0.42)     | 0.176                     |
| 5                 | 12.32    | 12.20                  | 0.12       | 0.014                     |
| 6                 | 13.14    | 12.83                  | 0.31       | 0.096                     |
| 7                 | 15.26    | 15.77                  | (0.51)     | 0.261                     |
| 8                 | 13.40    | 14.03                  | (0.63)     | 0.397                     |
| Total mean        |          |                        | 0.005      | 1.478                     |

<sup>A</sup>May be the mean of two reference values:

$$\text{RMSD} = (1.478/8)^{1/2}$$

$$= (0.185)^{1/2}$$

$$= 0.430$$

$$S_d = 0.460$$

<sup>B</sup>See Ref (2).

## REFERENCES

- (1) Grubbs, F. E., "On Estimating Precision of Measuring Instruments and Product Variability," *Journal of the American Statistical Association*, Vol 43, No. 2, June 1948, pp. 243–264.
- (2) Grubbs, F. E., "An Introduction to Some Precision and Accuracy of Measurement Problems," *Journal of Testing and Evaluation*, Vol 10, No. 4, July 1982, pp. 133–143.
- (3) Rose, C. D., "Methods for Assessing the Accuracy of On-Line Coal Analyzers," *Journal of Coal Quality*, Vol 10, No. 1, Jan.–Mar., 1991, pp. 19–28.
- (4) Jaech, J. L., "Further Tests of Significance for Grubbs Estimators," *Biometrics*, Vol 27, Dec. 1971, pp. 1097–1101.
- (5) Thompson, W. A., "The Problem of Negative Estimates of Variance Components," *Annals of Mathematical Statistics*, Vol 33, 1962, pp. 273–89.
- (6) Kalica, P. R. and Woodward R. C., "How Good is Grubbs?," *1993 Symposium on On-Line Analysis of Coal*, Vienna Austria.

*ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.*

*This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.*

*This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or [service@astm.org](mailto:service@astm.org) (e-mail); or through the ASTM website ([www.astm.org](http://www.astm.org)).*